

A survey of extremal combinatorics[†]

[†] McGill MATH 550 notes, Winter 2016.

Written by Eric Hanson,
based on lectures by Prof. Sergey Norin

April 2016

Combinatorics is the study of finite discrete structures[‡]. There are several subfields of combinatorics:

[‡] Say, collections of sets, or integers, or geometric sets.

- Enumerative combinatorics: try to understand how many objects exist with certain properties.
- Extremal combinatorics: what are the largest or smallest objects with certain properties. This is what we will focus on in this class.

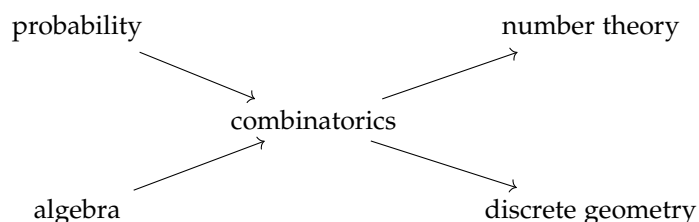


Figure 1: The flow of combinatorial arguments. Probabilistic and algebraic tools are used in combinatorial arguments, which yields results in number theory and discrete geometry.

Contents

1	<i>Set systems as an introduction to extremal combinatorial results</i>	2
1	<i>Representing Sets and Sperner Systems.</i>	5
2	<i>The Littlewood-Offord problem</i>	12
3	<i>Intersecting hypergraphs</i>	17
4	<i>Compression and Shadows</i>	23
5	<i>Turán type problems</i>	31
6	<i>Hypergraph Turán Problems</i>	42
6.1	<i>Turán density of Fano plane.</i>	45
7	<i>Ramsey Theory</i>	49
8	<i>Convexity</i>	60
9	<i>Incidence problems</i>	72
10	<i>Algebraic methods</i>	76
10.1	<i>Combinatorial Nullstellensatz (Alon and Tarsi 1989)</i>	76
10.2	<i>Kakeya needle problem (Kakeya 1917)</i>	81
10.3	<i>Shannon capacity (Shannon 1956)</i>	82

1 Set systems as an introduction to extremal combinatorial results

A *set system* is any $\mathcal{F} \subset \mathcal{P}(X)$. Let X be a finite set, and $\mathcal{P}(X)$ the collection of all subsets of X . Then $|\mathcal{P}(X)| = 2^{|X|}$. We'll denote $X^{(r)}$ as the collection of all r -element subsets of X .

Example. If $X = \{1, 2, 3\}$, then $\mathcal{P}(X) =$

$$\begin{array}{cccc} X^{(3)} & & \{1, 2, 3\} & \\ \\ X^{(2)} & \{1, 2\} & \{1, 3\} & \{2, 3\} \\ \\ X^{(1)} & \{1\} & \{2\} & \{3\} \\ \\ X^{(0)} & & \emptyset & \end{array}$$

Def. \mathcal{F} is an *intersecting set system* if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{F}$.

Largest intersecting set system

Given X with $|X| = n$, what is the maximum $|\mathcal{F}|$ with $\mathcal{F} \subset \mathcal{P}(X)$ intersecting? Let's start with $n = 3$ (returning to our example). We can select at least 4: every set that contains 1. This method gives us $2^{n-1} = |\mathcal{P}(X - \{1\})|^\dagger$. Why can't we do better? We can subdivide $\mathcal{P}(X)$ into pairs $\{Y, X - Y\}$. But $Y \cap (X - Y) = \emptyset$, so any intersecting \mathcal{F} contains at most one element of each pair. Thus, we can't do better than 2^{n-1} .

Largest intersecting set system with fixed number of elements

How large can \mathcal{F} intersecting be if $\mathcal{F} \subset X^{(r)}$ for some r ? Our answer will depend on r and $|X| = n$. Since $|X^{(r)}| = \binom{n}{r}$, that is an upper bound. When can we achieve it, i.e. when is $X^{(r)}$ intersecting? When $r > n/2$. In this case, every subset has more than half of the elements of X , so no two subsets can be disjoint.

If $r = n/2$, then because of complement pairing, the maximum is $\binom{n}{r}/2$. By choosing \mathcal{F} as all of the sets that contain a particular element, we get $\binom{n-1}{r-1}$. So for $r = n/2$, we have

$$\binom{n-1}{r-1} \leq \max |\mathcal{F}| \leq \frac{1}{2} \binom{n}{r}.$$

But for $r = n/2$, these two bounds are equal[‡]. For $r < n/2$, the answer is $\binom{n-1}{r-1}$; this is a lower bound by selecting all sets containing a given element. That this is an upper bound is the content of the following non-trivial theorem.

The following is a brief introduction to some of the results we will see in the course.

Note that the superscripted symbols [†], [‡], and ^{*} are reserved to point towards notes in this margin.

In this introduction, full references will be included in the margin; afterwards, they will appear at the end of each section.

In this case, we could also choose $X^{(2)} \cup X^{(3)}$, but that only works for n odd.

[†] since the number of subsets without 1 is in bijection with the number of subsets with 1. For a subset without 1, we add in 1, and get a new valid subset with 1. On the other hand, for a subset with 1, we take out 1 and get a valid subset without 1. These are clearly injective.

[‡] Why? the LHS is the number of sets containing 1. But half of the sets contain 1 and half don't, by the complement argument.

Theorem (Erdős-Ko-Rado). Let $r \leq n/2$. Let $\mathcal{F} \subset X^{(r)}$ with $|X| = n$ be intersecting. Then $|\mathcal{F}| \leq \binom{n-1}{r-1}$.

Next, let's consider other types of set systems.

Littlewood-Offord problem

We'll say \mathcal{F} is a *Sperner system* if $A \subset B$ for $A, B \in \mathcal{F}$, then $A = B$. Given X , $|X| = n$, what is $\max |\mathcal{F}|$ such that $\mathcal{F} \subset \mathcal{P}(X)$ is Sperner? Well, any $X^{(r)}$ is Sperner[†]. So we can achieve $\binom{n}{\lfloor n/2 \rfloor}$. It turns out that one cannot do better. For $n = 3$, we may make a graph by connecting subsets by edges. Then we need the largest independent set of this graph; it's easy to show in this case that this is $\binom{n}{\lfloor n/2 \rfloor}$. This is in fact the general result.

Problem (Littlewood-Offord 1938). Suppose we have non-zero numbers $a_1, a_2, \dots, a_n \in \mathbb{C}$. For $I \subset \{1, 2, \dots, n\}$, consider $\sum_{i \in I} a_i$. There are then 2^n possible such sums. What is the maximal number of them that can be equal to be zero?

Clearly we can shift to any number z instead of zero without changing the answer. Erdős solved this in 1945 for real numbers. Let's consider the case of positive numbers. First, if $a_1 = \dots = a_n = 1$, then the maximum number of equal sums is $\binom{n}{\lfloor n/2 \rfloor}$. Now for general positive numbers, consider $\mathcal{F}_z = \{I : \sum_{i \in I} a_i = z\}$. Then \mathcal{F}_z is a Sperner system, so by the previously quoted result its maximum size is $\binom{n}{\lfloor n/2 \rfloor}$. The Littlewood-Offord problem was treated in general in 1966 by Kleitman[‡].

Erdős-Szemerédi conjecture

Let $A \subset \mathbb{Z}_+$, $|A| = n$. Let $A + A = \{a + b : a, b \in A\}$. We wish to compare $|A + A|$ to $|A|$. Now, $\max |A + A| = \binom{n}{2} + n = \binom{n+1}{2}$ by choosing our elements so no two sums are the same unless they have to be*.

What is $\min |A + A|$? We'll choose a set with a lot of structure: $A = \{1, \dots, n\}$. Then $A + A = \{2, 3, \dots, 2n\}$, and $|A + A| = 2n - 1$. We'll prove that this is actually the minimum.

Proof. Let $a_1 < a_2 < \dots < a_n$. Then

$$\underbrace{a_1 + a_1 < a_1 + a_2 < \dots < a_1 + a_n}_n < \underbrace{a_2 + a_n < a_3 + a_n < \dots < a_n + a_n}_{n-1}$$

so we found $2n - 1$ different terms already. \square

P. Erdős, C. Ko, and R. Rado (1961). "Intersection Theorems For Systems Of Finite Sets". In: *The Quarterly Journal of Mathematics* 12.1, pp. 313–320. DOI: [10.1093/qmath/12.1.313](https://doi.org/10.1093/qmath/12.1.313)

[†] If you take any element out, you reduce the number of elements, so aren't in $X^{(r)}$ anymore.

J. E. Littlewood and A. C. Offord (1938). "On the Number of Real Roots of a Random Algebraic Equation". In: *Journal of the London Mathematical Society* s1-13.4, pp. 288–295. DOI: [10.1112/jlms/s1-13.4.288](https://doi.org/10.1112/jlms/s1-13.4.288)

[‡] D. J. Kleitman (1970). "On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors". In: *Advances in Mathematics* 5.1, pp. 155–157. ISSN: 0001-8708. DOI: [http://dx.doi.org/10.1016/0001-8708\(70\)90038-1](https://dx.doi.org/10.1016/0001-8708(70)90038-1).

* Intuitively then, we've achieved the maximum by a set with very little structure.

The elements are distinct because A is a set.

What about $|A \cdot A|$? Then $\max |A \cdot A| = \binom{n}{2} + n$, and $\min |A \cdot A| = 2n - 1$. The second is by $A = \{1, 2, 2^2, \dots, 2^{n-1}\}$. Generally we can transform additive problems into multiplicative by exponentiation, and back by the logarithm. Now, what's $\min(|A + A| + |A \cdot A|)$?

Conjecture (Erdős-Szemerédi).

$$|A + A| + |A \cdot A| = \Omega_\epsilon(|A|^{2-\epsilon})$$

That is, for all $\epsilon > 0$, there exists $c > 0$ such that for all $A \subset \mathbb{Z}_+ - \{0\}$,

$$|A + A| + |A \cdot A| \geq c_\epsilon |A|^{2-\epsilon}.$$

Elekes showed that $|A + A| + |A \cdot A| \geq c|A|^{5/4}$ (Elekes 1997) and Solymosi improved this to bound to $|A|^{4/3-\epsilon}$ (Solymosi 2009). In fact, this is related to the following geometric problem (see fig. 2 for an example).

Theorem (Szemerédi-Trotter theorem). *Suppose we have a collection L of lines in the plane, and P a collection of points in the plane. The set of incidences is $\{(p, \ell) : p \in P, \ell \in L, p \in \ell\}$. Let $I(P, L)$ denote the number of incidences for P and L^\dagger . Then*

$$I(P, L) \leq O(|P|^{2/3}|L|^{2/3} + |P| + |L|).$$

Elekes used this theorem to make progress on the Erdős-Szemerédi conjecture by mapping the set A to a collection of points and lines. He showed that if both $|A \cdot A|$ and $|A + A|$ are small, then you find too many incidences. The Szemerédi-Trotter theorem uses the following result:

Lemma (Crossing lemma). *Let G be a graph drawn in the plane with crossings. Suppose that G has m edges and n vertices. If $m \geq 4n$, then there are at least $\frac{m^3}{64n^2}$ crossings.*

From P and L , one makes a graph and applies the Crossing Lemma to prove Szemerédi-Trotter.

P. Erdős and E. Szemerédi (1983). "Studies in Pure Mathematics: To the Memory of Paul Turán". In: ed. by P. Erdős et al. Basel: Birkhäuser Basel. Chap. On sums and products of integers, pp. 213–218. ISBN: 978-3-0348-5438-2. DOI: [10.1007/978-3-0348-5438-2_19](https://doi.org/10.1007/978-3-0348-5438-2_19)

This qualitatively means the additive and multiplicative structures do not get along; you reduce $|A + A|$ only at the cost of increasing $|A \cdot A|$.

E. Szemerédi and W. T. Trotter (1983). "Extremal problems in discrete geometry". In: *Combinatorica* 3,3, pp. 381–392. ISSN: 1439-6912. DOI: [10.1007/BF02579194](https://doi.org/10.1007/BF02579194)

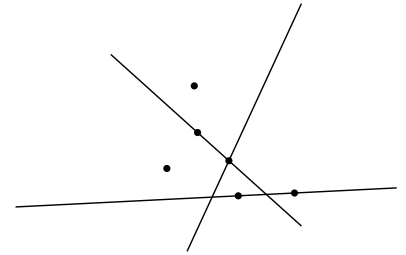


Figure 2: An example of a collection L of lines and P of points. We are interested in the number of incidences $I(P, L)$ between points in P and lines in L . Here, $I(P, L) = 5$.

[†] Clearly, $I(P, L) \leq |P| \cdot |L|$, as a subset of $P \times L$.

If $m > 3n - 6$, then there is at least one crossing, by the result stated at the beginning of the lecture, which comes from Euler's formula. In fact, this is how one proves the crossing lemma.

1 Representing Sets and Sperner Systems.

Let's start with some definitions.

Power set: Let X be a finite set with cardinality $|X| = n$. The *power set* is $\mathcal{P}(X)$, the collection of subsets of X . $|\mathcal{P}(X)| = 2^n$.

r -element subsets: The set $X^{(r)}$ is the collection of all r -element subsets of X . The cardinality $|X^{(r)}| = \binom{n}{r}$.

Set system: A *set system* $\mathcal{F} \subset \mathcal{P}(X)$ on X is a collection of subsets of X . For example, $\mathcal{F} = \{\emptyset, \{1\}, \{1, 2\}, \{2, 3\}, \{3\}\}$ is a set system on $\{1, 2, 3\}$.

r -graph: If $\mathcal{F} \subset X^{(r)}$, then we call \mathcal{F} an r -graph or *hypergraph*. 2-graphs are just ordinary graphs: $\mathcal{F} \subset X^{(2)}$ can be thought of as a graph with vertex set X , edge set \mathcal{F} .

We'll also frequently use the notation $[n] = \{1, 2, \dots, n\}$.

NOW, GIVEN A SET SYSTEM $\{A_1, A_2, \dots, A_m\}$, we want to “reduce” the sets such that distinct sets remain distinct. That is, we want to find $S \subset X$ as small as possible such that $\{A_1 \cap S, A_2 \cap S, \dots, A_m \cap S\}$ are all distinct. Let's start with two sets, $\{A_1, A_2\}$. We want S as small as possible such that $A_1 \cap S \neq A_2 \cap S$. In this case, we can simply choose S to be a singleton of an element which is in one set but not the other[†]. If we use our example earlier, $\{\emptyset, \{1\}, \{1, 2\}, \{2, 3\}, \{3\}\}$ on $[3]$, we cannot remove any; $|\mathcal{P}([2])| = 4$ and we have five elements. If our set system is $\{\emptyset, \{1, 2\}, \{2, 3\}, \{3\}\}$ on $[3]$, then we may remove 1; that is, $S = \{2, 3\}$. This motivates a question: How small can S be as a function of m ?

Theorem 1.1. *Let $\mathcal{F} = \{A_1, A_2, \dots, A_m\}$ be a set system. Then there exists S , $|S| \leq m - 1$ such that $A_1 \cap S, A_2 \cap S, \dots, A_m \cap S$ are all distinct.*

Remark. The bound is tight: the set system $\{\emptyset, \{1\}, \{2\}, \dots, \{m - 1\}\}$ has the property that if we remove any element, we collapse two sets to the empty set.

Proof. Choose S as small as possible such that $A_1 \cap S, A_2 \cap S, \dots, A_m \cap S$ are all distinct, and assume that $|S| \geq m$. Let $A'_i = A_i \cap S$. By the minimality of S for every $x \in S$ there exists $i, j \in [m]$ such that $A'_i - \{x\} = A'_j - \{x\}$, with $i \neq j$.

Now, construct a graph on the vertex set $[m]$ as follows. For each $x \in S$, choose one pair i and j ($i \neq j$) such that $A'_i - \{x\} = A'_j - \{x\}$, and join i and j by an edge[‡]. This graph has m vertices and $|S| \geq m$ edges, so it contains a cycle*. Without loss of generality, assume there is a cycle on vertices $1, 2, \dots, k$ in order. Then there exists distinct

We begin our systematic investigation into extremal combinatorics.

Note that S acts by deleting elements from our base set X .

[†] which always exists because $A_1 \neq A_2$.

[‡] $A'_i - \{x\} = A'_j - \{x\}$ is equivalent to $A_i \Delta A_j = \{x\}$, where the symmetric difference $X \Delta Y = (X \cup Y) - (X \cap Y)$. Because of this, we will make a new edge each time: if $x, y \in S$ yielded the same edge, then $\{x\} = A_i \Delta A_j = \{y\}$.

* Easy to see by picture; draw $m - 1$ edges on m vertices, and then if you don't have a cycle yet, you have a line, and no matter how you place the last edge, you get a cycle.

$x_1, x_2, \dots, x_k \in S$ such that $A_1 \triangle A_2 = \{x_1\}$, $A_2 \triangle A_3 = \{x_2\}$, \dots , $A_{k-1} \triangle A_k = \{x_{k-1}\}$, and $A_k \triangle A_1 = \{x_k\}$.

We can take the symmetric difference of all of them:

$$\emptyset = (A_1 \triangle A_2) \triangle (A_2 \triangle A_3) \triangle \dots \triangle (A_k \triangle A_1) = \{x_1, x_2, \dots, x_k\}$$

On the left, we have two of each set, so we can regroup and commute to obtain the empty set, using $A \triangle A = \emptyset$. On the right, we have the symmetric difference of distinct singletons, which is just the union. This is a contradiction, so our minimal S must have $|S| \leq m - 1$. \square

Before we continue finding ways to represent sets, we'll need some graph theoretic tools. First, some definitions.

Bipartite: A graph G is *bipartite* with bipartition (V_1, V_2) if every edge of G contains one vertex of V_1 and one vertex of V_2 .

Matching: A collection of edges M of G is a *matching* of V_1 into V_2 if for every $v \in V_1$, M contains exactly one edge containing v , and for $v \in V_2$, at most one edge. This is illustrated in fig. 3.

Neighborhood: For $S \subset V_1$, define the *neighborhood* $N(S)$ as the set of vertices adjacent to at least one vertex in S .

The following result[†] connects these ideas.

Theorem 1.2 (Hall's marriage theorem). *Let G be a bipartite graph with bipartition (V_1, V_2) . Then G contains a matching of V_1 into V_2 if and only if*

$$|N(S)| \geq |S| \text{ for every } S \subset V_1. \quad (\text{Hall's condition})$$

Proof. The condition is necessary because you need to have enough vertices available in $N(S)$ for elements of S to match into. We will prove sufficiency by induction on $|V_1|$. The base case is immediate. For the induction step, we will split into two cases.

Case 1: For every $S \subset V_1$ with $S \neq \emptyset$ and $S \neq V_1$, we have that $|N(S)| > |S|$. In this case, choose $v \in V_1$; then v has a $w \in V_2$ adjacent to it, because $|N(\{v\})| > |\{v\}| = 1$. Apply the induction hypothesis to $G - \{v, w\}$.

We then just need to check Hall's condition on $G' = G - \{v, w\}$. For every $S \subset V_1 - \{v\}$, we have

$$|N'(S)| \geq |N(S)| - 1 \geq |S|,$$

where N' is the neighborhood with respect to G' . The first inequality holds because we removed at most one neighbor

Note the symmetric difference is commutative and associative.

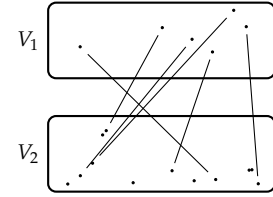


Figure 3: Think of elements of V_1 as job applicants, and V_2 as positions. Then a matching of V_1 into V_2 is an arrangement so that every job applicant has a position, but some positions could be unfilled.

[†] Hall 1935

by removing w . The second inequality holds from our assumption in this case. Then the induction hypothesis yields a matching on $V_1 - \{v\}$ into $V_2 - \{w\}$, which we can extend to a matching on V_1 into V_2 by matching v to w .

Case 2: There exists $S \subset V_1$ with $S \neq \emptyset$ and $S \neq V_1$, such that $|N(S)| = |S|$. By induction hypothesis, there exists a matching M_1 of S into $N(S)$.

It remains to find a matching from $V_1 - S$ into $V_2 - N(S)$. By induction hypothesis, it is enough to show that for every $T \subset V_1 - S$, we have

$$|N(T) \cap (V_2 - N(S))| \geq |T|.$$

Since $S \cup T \subset V_1$, by assumption, we have Hall's condition

$$|N(S \cup T)| \geq |S \cup T| = |S| + |T|.$$

We know $N(S \cup T) = N(S) \cup N(T) = N(S) \cup (N(T) - N(S))$. So $|N(S \cup T)| = |N(S)| + |N(T) - N(S)|$. So Hall's condition becomes

$$|N(T) - N(S)| \geq |T|$$

as desired. \square

Let's employ Hall's theorem to represent sets. Let

$$\mathcal{F} = \{A_1, A_2, \dots, A_m\}$$

be a set system. A *system of distinct representatives* for \mathcal{F} is a collection $\{x_1, \dots, x_m\}$ of elements such that x_1, \dots, x_m are pairwise distinct, and $x_i \in A_i$ for $i \in [m]$. Given \mathcal{F} , one can consider the bipartite graph G with bipartition (V_1, V_2) such that $V_1 = \mathcal{F}$ and $V_2 = \bigcup_{i \in [m]} A_i$. We join A_i to x iff $x \in A_i$. Then a system of distinct representatives for \mathcal{F} is exactly a matching on G from V_1 into V_2 . Hall's theorem then immediately implies the following result.

Corollary 1.3. *A set system $\mathcal{F} = \{A_1, \dots, A_m\}$ has a system of distinct representatives if and only if for every $\mathcal{F}' \subset \mathcal{F}$,*

$$|\mathcal{F}'| \leq \left| \bigcup_{A \in \mathcal{F}'} A \right|.$$

GIVEN A SET X , there is a natural bipartite graph and matching which will prove useful.

Corollary 1.4. *Let X be a set with $|X| = n$. Let G be a bipartite graph with bipartition $(X^{(r)}, X^{(r-1)})$ such that $A \in X^{(r)}$ is adjacent to $B \in X^{(r-1)}$ if $B \subset A$. Then if $r > n/2$, the graph G has a matching of $X^{(r)}$ into $X^{(r-1)}$.*

Here's the trick.

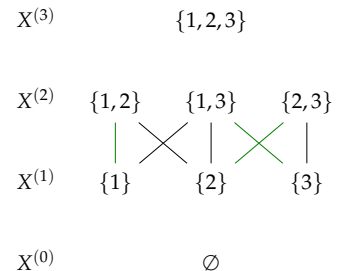


Figure 4: Example of the graph relation on $G = (X^{(2)}, X^{(3)})$, with a matching highlighted in green.

Remark. This corollary implicitly shows $\binom{n}{r} \leq \binom{n}{r-1}$ if $r > n/2$.

Proof. It suffices to check [Hall's condition](#). For every $\mathcal{A} \subset X^{(r)}$, we want $|N(\mathcal{A})| \geq |\mathcal{A}|$. Label

$$\mathcal{B} := N(\mathcal{A}) = \{B \in X^{(r-1)} : B \subset A, \text{ for some } A \in \mathcal{A}\}.$$

Every element of $X^{(r)}$ is incident to r edges of G^\dagger . So we have $|\mathcal{A}|r$ edges leaving \mathcal{A} , ending in \mathcal{B} . Every element of $X^{(r-1)}$ is incident to $n - r + 1$ edges of G , which can be seen by the fact that there are $n - (r - 1)$ possible elements to add to a set $B \in X^{(r-1)}$ to obtain a superset in $X^{(r)}$. So we have at most $|\mathcal{B}|(n - r + 1)$ edges leaving \mathcal{B} , ending in \mathcal{A}^\ddagger . So $|\mathcal{A}|r \leq |\mathcal{B}|(n - r + 1)$. But by assumption $r \geq (n - r + 1)$, so $|\mathcal{B}| \geq |\mathcal{A}|$ as desired. \square

Recall that $\mathcal{F} \subset \mathcal{P}(X)$ is a *Sperner system* if for all $A, B \in \mathcal{F}$, if $A \leq B$, then $A = B$. We wish to find $\max |\mathcal{F}|$ such that \mathcal{F} is a Sperner system, as a function $|X| = n$. Note that $X^{(r)}$ is always Sperner, and $|X^{(r)}| = \binom{|X|}{r}$, which is maximized when $r = \lfloor n/2 \rfloor$.

Theorem 1.5 (Sperner 1928). *If $\mathcal{F} \subset \mathcal{P}(X)$ is Sperner, then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$, where $n = |X|$.*

Proof. An ordered collection (A_1, A_2, \dots, A_k) of sets in $\mathcal{P}(X)$ is a *chain* if $A_1 \subsetneq A_2 \subsetneq A_3 \subsetneq \dots \subsetneq A_k$.

It is enough to show that $\mathcal{P}(X)$ can be partitioned into $\binom{n}{\lfloor n/2 \rfloor}$ chains. Indeed, every Sperner system can contain ≤ 1 element from each chain in the partition; see [fig. 5](#) for an example. Better yet, we partition $\mathcal{P}(X)$ into chains such that every chain contains an element of $X^{(\lfloor n/2 \rfloor)}$.

Let's begin by partitioning all subsets of X of size $\geq \lfloor n/2 \rfloor$. We will first do this inductively starting from $X^{(\lfloor n/2 \rfloor)}$, and extending the partition to $X^{(\lfloor n/2 \rfloor)} \cup X^{(\lfloor n/2 \rfloor + 1)} \cup \dots \cup X^{(k)}$ to $X^{(k+1)}$ using the matching obtained in [corollary 1.4](#) from $X^{(k)}$ to $X^{(k+1)}$ by adding each element of $X^{(k+1)}$ to the chain of the set it's matched to. Note that are chains are not maximal; some (all but one) truncate before they reach the top, $X^{(n)}$. Then we can extend the partition to sets of size $< \lfloor n/2 \rfloor$ by symmetry. \square

Remark. This proof is instructive and provides the useful technique of partitioning into chains. But we can prove stronger results with slicker proofs.

Suppose $k < n/2$ and we want to find the maximum size Sperner system such that every set in the system has size $\leq k$. As one may guess, the maximum size will be $|X^{(k)}| = \binom{n}{k}$. To show this, we'll use the following result.

† Each edge corresponds to taking an element away from the set.

‡ Since not every edge leaving \mathcal{B} needs to reach something in \mathcal{A} (it could reach something in $X^{(r)} - \mathcal{A}$), it is only "at most."

A Sperner system is a system of sets such that no two are comparable. A dual notion is a system of sets such that all are comparable.

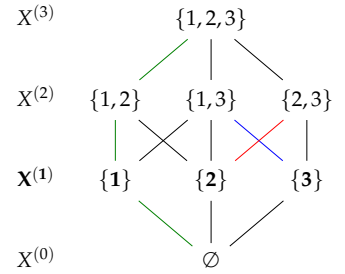


Figure 5: Consider $X^{(1)}$, a natural maximal Sperner system. Note that each element of $X^{(1)}$ can form a distinct chain, such as $\emptyset \subsetneq \{1\} \subsetneq \{1,2\} \subsetneq \{1,2,3\}$.

Theorem 1.6 (Lubell, Meshalkin, Yamamoto, Bollobás, and possibly others, A.K.A. the LYM inequality). Let $\mathcal{F} \subset \mathcal{P}(X)$, $|X| = n$ be a Sperner system. Let $\mathcal{F}_k = \mathcal{F} \cap X^{(k)}$ be the set of k element sets in \mathcal{F} , and let $f_k = |\mathcal{F}_k|$. Then

$$\sum_{k=0}^n \frac{f_k}{\binom{n}{k}} \leq 1. \quad (\text{LYM})$$

Remark. We have

$$1 \geq \sum_{k=0}^n \frac{f_k}{\binom{n}{k}} \geq \sum_{k=0}^n \frac{f_k}{\binom{n}{\lfloor n/2 \rfloor}},$$

thus

$$\binom{n}{\lfloor n/2 \rfloor} \geq \sum_{k=0}^n f_k = |\mathcal{F}|$$

which is Sperner's theorem.

Proof. Let's assume $X = [n]$ for convenience. Consider all maximum chains in $\mathcal{P}(X)$ and count how many chains an element of \mathcal{F} belongs to. Each of the maximal chains is of the form $\emptyset = A_0, A_1, \dots, A_n = X = [n]$, and $|A_i| = i$. So each maximal chain corresponds to an ordering a_1, a_2, \dots, a_n of $[n]$, where $\{a_i\} = A_i - A_{i-1}$, is the element you add to A_i to get the next set in the chain.

Thus, there are $n!$ maximal chains (the number of re-orderings of $[n]$). Consider $F \in X^{(k)}$. How many maximal chains is F in? If $k = 0, n$, F is in every chain, so $n!$. If $k = 1$, then F is in $(n-1)!$ chains. If $k = 2$, then $2!(n-2)!$. In general, F is in $k!(n-k)!$ maximal chains[†]. Each maximal chain contains ≤ 1 element of \mathcal{F} . The total number of elements of \mathcal{F} in all maximum chains is

$$\sum_{k=0}^n f_k k!(n-k)! \leq n!$$

Dividing by $n!$, we obtain the LYM inequality. \square

Remark. Let's consider an alternate proof. Let C be a uniformly randomly chosen maximal chain, and consider the expectation value of the number of elements of $C \cap \mathcal{F}$. Of course, there is at most 1 element, since \mathcal{F} is a Sperner system. On the other hand,

$$\begin{aligned} \mathbb{E}(|C \cap \mathcal{F}|) &= \sum_{k=0}^n \mathbb{E}(|C \cap \mathcal{F}_k|) = \sum_{k=0}^n f_k \cdot (\text{probability that a set of size } k \text{ is in } C) \\ &= \sum_{k=0}^n f_k \cdot \frac{1}{(\text{number of sets of size } k)} = \sum_{k=0}^n \frac{f_k}{\binom{n}{k}} \leq 1. \end{aligned}$$

When does equality hold in LYM? Certainly when $\mathcal{F} = X^{(r)}$ for any r . We'd like to show this condition is necessary as well, but to do so, we'll first prove a more refined inequality in which equality is

Bollobás 1965; Lubell 1966; Meshalkin 1963; Yamamoto 1954

[†] We choose k to get to the set, then $(n-k)$ to finish the chain.

easier to check. Then we'll use this to show sufficiency for equality in the LYM inequality.

Theorem 1.7 (Local LYM inequality). *Let $\mathcal{A} \subset X^{(r)}$, and $|X| = n$. Define $\partial\mathcal{A} \subset X^{(r-1)}$ the shadow of \mathcal{A} by*

$$\partial\mathcal{A} := \{B \in X^{(r-1)} : B \supseteq A \text{ for some } A \in \mathcal{A}\}.$$

Then

$$\frac{|\partial\mathcal{A}|}{\binom{n}{r-1}} \geq \frac{|\mathcal{A}|}{\binom{n}{r}} \quad (\text{Local LYM})$$

or equivalently,

$$r|\mathcal{A}| \leq |\partial\mathcal{A}|(n-r+1).$$

Moreover, equality holds if and only if $\mathcal{A} = \emptyset$ or $\mathcal{A} = X^{(r)}$.

Proof.

$$r|\mathcal{A}| = |\{(B, A) : B \in \partial\mathcal{A}, A \in \mathcal{A}, B \subset A\}| \leq |\partial\mathcal{A}|(n-r+1)$$

as seen in corollary 1.4. If equality holds, then \mathcal{A} contains all supersets in $X^{(r)}$ of all sets in $\partial\mathcal{A}$.

Consider the graph as in corollary 1.4: there are no edges from $A \cup \partial\mathcal{A}$ to the remaining vertices, so since G is connected[†], we have equality in (Local LYM). \square

Theorem 1.8. *The equality in the LYM inequality (LYM) holds iff $\mathcal{F} = X^{(r)}$ for some r .*

Proof. Inductively define $G_n = \mathcal{F}_n$, and for $k < n$, $G_k = \partial G_{k+1} \cup \mathcal{F}_k$.

Let $\phi_k = \frac{f_k}{\binom{n}{k}}$ be the proportion of sets of \mathcal{F} in $X^{(k)}$. Similarly, set $\gamma_k = \frac{|G_k|}{\binom{n}{k}}$. By the local LYM,

$$\frac{|\partial G_{k+1}|}{\binom{n}{k}} \geq \frac{|G_{k+1}|}{\binom{n}{k+1}} = \gamma_{k+1}$$

So,

$$\gamma_k = \frac{|\partial G_{k+1} \cup \mathcal{F}_k|}{\binom{n}{k}} = \frac{|\partial G_{k+1}|}{\binom{n}{k}} + \frac{|\mathcal{F}_k|}{\binom{n}{k}} \geq \gamma_{k+1} + \phi_k$$

where we are using that \mathcal{F} is Sperner, so that ∂G_{k+1} is disjoint from \mathcal{F}_k . Thus, we have $\gamma_k \geq \gamma_{k+1} + \phi_k$, with equality iff $\gamma_{k+1} = 1$ or $\gamma_{k+1} = 0$.

Note $\gamma_n = \phi_n$. Then

$$\gamma_{n-1} \geq \gamma_n + \phi_{n-1} = \phi_n + \phi_{n-1}$$

$$\gamma_{n-2} \geq \gamma_{n-1} + \phi_{n-2} = \phi_n + \phi_{n-1} + \phi_{n-2}$$

etc.

$$\gamma_k \geq \phi_n + \phi_{n-1} + \cdots + \phi_k.$$

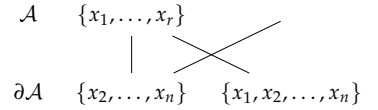


Figure 6: If \mathcal{A} contains all supersets in the $X^{(r)}$ layer of sets in the shadow $\partial\mathcal{A}$, then as long as $\mathcal{A} \neq \emptyset$, it must contain every set; here, for example, \mathcal{A} has to include the endpoint of the edge leaving the bottom left vertex.

[†] as is easy to check

G_k is the set of all k -element sets which are subsets of sets in \mathcal{F} .

Since that is when we have equality in eq. (Local LYM).

Hence, $1 \geq \gamma_0 \geq \phi_n + \phi_{n-1} + \cdots + \phi_0$. This is the LYM inequality. But equality holds if $\gamma_k = \gamma_{k+1} + \phi_k$ for each k , i.e. $\gamma_{k+1} = 1$ or $\gamma_{k+1} = 0$ for each k . Assuming we have equality, we use that γ_k is non-increasing with k , so there must exist k_0 such that $\gamma_{k_0} = 1$, and $\gamma_{k_0+1} = 0$ (writing $\gamma_{n+1} = 0$). Thus, there are no sets in \mathcal{F} of size at least $k+1$. In other words, $G_{k+1} = \emptyset$, and $G_k = \mathcal{F}_k = X^{(k)}$. But then we must have $\mathcal{F} = X^{(k)}$ as desired, since \mathcal{F} may not have any super sets or subsets of $X^{(k)}$, i.e., any other set in $\mathcal{P}(X)$. \square

Remark. The matchings with $X^{(r)}$ and $X^{(r-1)}$ are the essential objects here in proving the local LYM, and hence LYM and its equality.

Exercise. Prove Sperner's theorem using the original way: partitioning into $\binom{n}{\lfloor n/2 \rfloor}$ chains by induction on n instead of Hall's theorem.

This is theorem 2.2.

References for Section 1.

- Bollobás, B. (1965). "On generalized graphs". In: *Acta Mathematica Academiae Scientiarum Hungarica* 16.3, pp. 447–452. ISSN: 1588-2632. DOI: [10.1007/BF01904851](https://doi.org/10.1007/BF01904851) (cit. on p. 9).
- Hall, P. (1935). "On Representatives of Subsets". In: *Journal of the London Mathematical Society* s1-10.1, pp. 26–30. DOI: [10.1112/jlms/s1-10.37.26](https://doi.org/10.1112/jlms/s1-10.37.26) (cit. on p. 6).
- Lubell, D. (1966). "A short proof of Sperner's lemma". In: *Journal of Combinatorial Theory* 1.2, pp. 299–. ISSN: 0021-9800. DOI: [http://dx.doi.org/10.1016/S0021-9800\(66\)80035-2](http://dx.doi.org/10.1016/S0021-9800(66)80035-2) (cit. on p. 9).
- Meshalkin, L. D. (1963). "Generalization of Sperner's Theorem on the Number of Subsets of a Finite Set". In: *Theory of Probability & Its Applications* 8.2, pp. 203–204. DOI: [10.1137/1108023](https://doi.org/10.1137/1108023) (cit. on p. 9).
- Sperner, E. (1928). "Ein Satz über Untermengen einer endlichen Menge". In: *Mathematische Zeitschrift* 27.1, pp. 544–548. ISSN: 1432-1823. DOI: [10.1007/BF01171114](https://doi.org/10.1007/BF01171114) (cit. on p. 8).
- Yamamoto, K. (Dec. 1954). "Logarithmic order of free distributive lattice". In: *J. Math. Soc. Japan* 6.3-4, pp. 343–353. DOI: [10.2969/jmsj/00630343](https://doi.org/10.2969/jmsj/00630343) (cit. on p. 9).

2 The Littlewood-Offord problem

Problem (Littlewood and Offord 1938). Given (z_1, \dots, z_n) complex numbers, with $|z_i| \geq 1$. Consider the 2^n possible sums formed by the z_i 's. How many sums can have pairwise distances less than 1 from each other? If $z_1 = z_2 = \dots = z_n = 1$, then we get $\binom{n}{\lfloor n/2 \rfloor}$ equal sums.

Given (z_1, \dots, z_n) , and subset of indices $A \subset [n]$, let $Z_A = \sum_{i \in A} z_i$, where we define $Z_\emptyset = 0$ and $Z_{\{i\}} = z_i$. We say $\mathcal{F} \subset \mathcal{P}([n])$ is *Littlewood-Offord* (LO) with respect to (z_1, \dots, z_n) if $|Z_A - Z_B| < 1$ for all $A, B \in \mathcal{F}$. In this notation, we are looking for the largest collection \mathcal{F} such that \mathcal{F} is LO with respect to (z_1, \dots, z_n) .

Conjecture. If $(z_1, \dots, z_n) \in \mathbb{C}^n$ with $|z_i| \geq 1$, and \mathcal{F} is LO with respect to (z_1, \dots, z_n) , then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

Remark. Littlewood and Offord were not combinatorialists; instead they were interested in studying the roots of random polynomials.

Theorem 2.1 (Erdős 1945). If x_1, \dots, x_n are real, $|x_i| \geq 1$ for every i , and \mathcal{F} is LO with respect to (x_1, \dots, x_n) , then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

Proof. We'll change notation to X_A instead of Z_A for this real case. If $x_1, \dots, x_n \geq 1$, then \mathcal{F} is Sperner. Suppose $A \subsetneq B$; then

$$|X_A - X_B| = X_B - X_A = X_{B-A} \geq |B - A| \geq 1. \quad \checkmark$$

The general problem for reals can be reduced to this case of $x_i > 0$ as follows. Suppose that \mathcal{F} is LO with respect to (x_1, \dots, x_n) . Then we will construct a family \mathcal{F}' which is LO with respect to

$(x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n)$ with $|\mathcal{F}'| = |\mathcal{F}|$. This will suffice to prove the theorem.

Let $\mathcal{F}' = \{A \triangle \{i\} : A \in \mathcal{F}\}^\dagger$. Clearly $|\mathcal{F}'| = |\mathcal{F}|$. We've reduced all sums by x_i , as shown by the following. If $i \in A$, then $X'_{A \triangle \{i\}} = \left(\sum_{j \in A} x_j\right) - x_i$. If $i \notin A$, then $X'_{A \triangle \{i\}} = \left(\sum_{j \in B} x_j\right) - x_i$. Hence, all sums are still within 1 of each other, completing the proof. \square

[†] We remove i if it's there, and add it otherwise.

Def. A chain $A_1 \subset A_2 \subset \dots \subset A_k$ in $\mathcal{P}([n])$ is *symmetric* if $|A_{i+1}| = |A_i| + 1$ for $i = 1, \dots, k-1$. Additionally, we require $|A_1| + |A_k| = n$.

Note that in particular, a symmetric chain intersects $[n]^{(\lfloor n/2 \rfloor)}$. An example of symmetric chains is shown in fig. 7.

Our previous method, when proving Sperner's theorem, didn't actually guarantee symmetric chains; see fig. 7. We will do this now, which provides a new proof of Sperner's theorem.

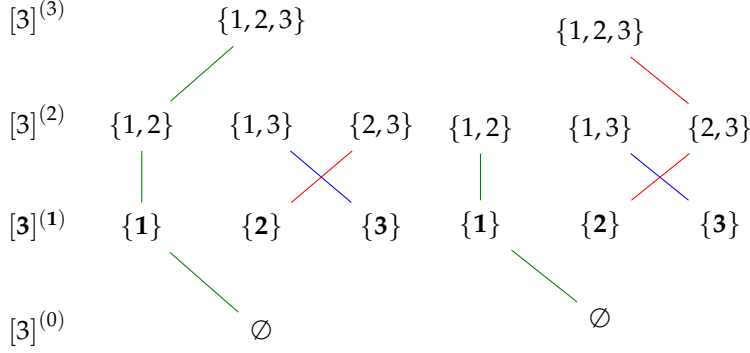


Figure 7: *Left:* We draw our favorite $\mathcal{P}([n])$, partitioned into symmetric chains, which thus each intersect $[n]^{(\lfloor n/2 \rfloor)}$. *Right:* A partition of $\mathcal{P}([n])$ into chains which intersect $[n]^{(\lfloor n/2 \rfloor)}$, which we could obtain from the proof of theorem 1.5. But these chains are not symmetric.

Theorem 2.2. $\mathcal{P}([n])$ can be partitioned into symmetric chains.

Proof by induction on n . The base case $n = 1$ is immediate. Induction step: Let C_1, C_2, \dots, C_L be symmetric chains forming a partition of $\mathcal{P}[n-1]$. Consider $C_i = (A_1, A_2, \dots, A_k)$. Form

$$C'_i = (A_1, A_2, \dots, A_k, A_k \cup \{n\})$$

$$C''_i = (A_1 \cup \{n\}, A_2 \cup \{n\}, \dots, A_{k-1} \cup \{n\}).$$

We can easily see that given that C_i was a symmetric chain on

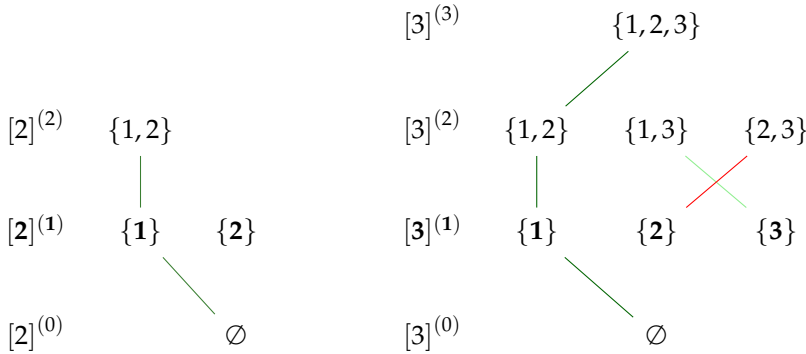


Figure 8: An illustration of the induction step, from $n = 2$ on the left to $n = 3$ on the right. On the left, two chains: $C_1 = (\emptyset, \{1\}, \{1,2\})$ in green, and $C_2 = (\{2\})$. We create $C'_1 = (\emptyset, \{1\}, \{1,2\}, \{1,2,3\})$ in dark green, and $C''_1 = (\{3\}, \{1,3\})$ in light green. We also create $C'_2 = (\{2\}, \{2,3\})$ in red, and $C'_3 = (\emptyset)$, an empty chain.

$\mathcal{P}([n-1])$, we have C'_i and C''_i are both symmetric chains on $\mathcal{P}([n])$. Moreover,

$$\{C'_1, \dots, C'_L\} \cup \{C''_1, \dots, C''_L\}$$

form a partition of $\mathcal{P}([n])$, as follows: For any set $A \in \mathcal{P}([n])$, if $n \notin A$, then $A \in C_i$ for some i , and hence $A \in C'_i$. If $n \in A$, then $A - \{n\} \in C_i = (A_1, A_2, \dots, A_k)$ for some i , so $A = A_j$ for some j . If $j = k$, then $A \in C'_i$, otherwise $A \in C''_i$. In any case, if $A \in \mathcal{P}([n])$, A is in one of these chains. Finally, A cannot be in two chains, otherwise we have that $\{C_i\}$ was not a partition of $\mathcal{P}([n-1])$. \square

Any partition of $\mathcal{P}([n])$ symmetric chains has $\binom{n}{\lfloor n/2 \rfloor}$ sets, because each chain must intersect with one point in $[n]^{(\lfloor n/2 \rfloor)}$. In the proof,

we took a partition into $\binom{n-1}{\lfloor (n-1)/2 \rfloor}$ chains, and seem to have created $2\binom{n-1}{\lfloor (n-1)/2 \rfloor} \neq \binom{n}{\lfloor n/2 \rfloor}$ chains. The catch is that C_i'' is an empty chain if $k = 1$.

LET US COUNT the sizes of symmetric chains. Suppose C_1, \dots, C_L is a partition of $\mathcal{P}([n])$ into symmetric chains, where $L = \binom{n}{\lfloor n/2 \rfloor}$. How many chains are there of length $r = n + 1$ in the partition? Exactly one, the chain which contains the empty set and must therefore contain the set of n elements.

What about $r = n$? Zero.

What about $r = n - 1$? $n - 1$ chains, because it must start the collection of 1 element sets and go to the collection of $n - 1$ element sets. There are n 1-element sets, but one is already in the maximal chain, so we are left with $n - 1$.

What about $r = (n + 1) - 2i$? These are the symmetric chains which start at the i th level $[n]^{(i)}$ and go to the $n - i$ th level $[n]^{(n-i)}$. The result is $\binom{n}{i} - \binom{n}{i-1}$. That is because there are $\binom{n}{i}$ elements in $[n]^{(i)}$, but $\binom{n}{i-1}$ of them are part of longer chains, the number of which is the number of elements in the level $[n]^{(i-1)}$.

Let us formulate analogues of these ideas to solve the Littlewood-Offord problem in \mathbb{R}^d .

Theorem 2.3 (Kleitman 1970). *Let (x_1, \dots, x_n) be vectors in \mathbb{R}^d , with $\|x_i\| \geq 1$. As before, define for $A \subset [n]$, $X_A = \sum_{i \in A} x_i$, and say that $\mathcal{F} \subset \mathcal{P}([n])$ is LO with respect to (x_1, \dots, x_n) if for each $A, B \in \mathcal{F}$, we have $\|X_A - X_B\| < 1$. Let \mathcal{F} be LO with respect to (x_1, \dots, x_n) . Then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

Proof. We will say $\mathcal{C} \subset \mathcal{P}([n])$ is *sparse* (with respect to (x_1, \dots, x_n)) if $\|X_A - X_B\| > 1$ for all $A, B \in \mathcal{C}$ with $A \neq B$. It is enough to show that $\mathcal{P}([n])$ can be partitioned into $\binom{n}{\lfloor n/2 \rfloor}$ sparse sets[†].

We say that a partition C_1, \dots, C_L of $\mathcal{P}([n])$ is *symmetric* if it contains exactly $\binom{n}{i} - \binom{n}{i-1}$ “chains” of order $(n + 1) - 2i$ for $i = 0, 1, \dots, \lfloor n/2 \rfloor$, and no chains with sizes not congruent to $n + 1 \pmod{2}$. In particular, $L = \binom{n}{\lfloor n/2 \rfloor}$.

We will show that $\mathcal{P}[n]$ has a symmetric partition into sparse sets. Then we will have found $L = \binom{n}{\lfloor n/2 \rfloor}$ sparse sets, completing the proof. Let’s proceed by induction on n . For $n = 1$, we have $X_\emptyset = 0$, and $X_{\{1\}} = x_1$. Then $\{\emptyset, \{1\}\}$ is sparse if $\|X_\emptyset - X_{\{1\}}\| > 1$, which holds because $\|x_1\| > 1$. Induction step: Let C_1, C_2, \dots, C_L be a symmetric partition of $\mathcal{P}([n - 1])$ into sets sparse with respect to (x_1, \dots, x_{n-1}) .

[†] A LO family may only contain one element of each sparse set.

Here, “chain” simply means a set C_i for some $i \in [L]$.

Let $C_i = \{A_1, \dots, A_k\}$. We'd like to form sparse sets

$$C'_i = \{A_1, A_2, \dots, A_k, A_k \cup \{n\}\}$$

$$C''_i = \{A_1 \cup \{n\}, A_2 \cup \{n\}, \dots, A_{k-1} \cup \{n\}\}.$$

Our sets A_1, \dots, A_k do not have an ordering this time, and we may choose any set to be A_k . We will have to use this freedom; consider translating the sums in some chain $C_i = (A_1, \dots, A_k)$ by adding the vector x_n . We want C'_i to be sparse, but then we need $X_{A_k \cup \{n\}}$ far from each X_{A_j} , which does not always need to happen, as illustrated in fig. 9.

Assume WLOG that $x_n = (\alpha, 0, 0, \dots, 0)$ only has non-zero first coordinate with $\alpha > 0$. Let $p(v)$ denote the first coordinate of v . Then p is a linear function, and $p(v) \leq \|v\|$ for each v .

Let A_k in the construction above be chosen so that $p(X_{A_k}) \geq p(X_{A_j})$ for each j [†]. Then to show that C'_i is sparse, it suffices to see that

$$\begin{aligned} \|X_{A_k \cup \{n\}} - X_{A_j}\| &\geq p(X_{A_k \cup \{n\}} - X_{A_j}) \\ &= p(x_{A_k}) + p(x_n) - p(X_{A_j}) \geq p(x_n) \geq 1. \end{aligned}$$

Hence, we follow the same procedure as in theorem 2.2 to conclude that indeed we did produce a symmetric partition. \square

LET US NOW CONSIDER $\mathbb{Z}_p := \mathbb{Z}/p\mathbb{Z}$, the integers modulo a prime p . Let $x_1, \dots, x_n \in \mathbb{Z}_p - \{0\}$. We will say $\mathcal{C} \subset \mathcal{P}([n])$ is *sparse* with respect to x_1, \dots, x_n if $X_A \neq X_B$ for all $A, B \in \mathcal{C}$ with $A \neq B$. Let us define

$$\Sigma(z) := |\{A \subset \mathcal{P}([n]) : X_A = z\}|.$$

We are interested in $\max \Sigma(z)$. First, if each $x_1 = \dots = x_n = 1$, then we see that may achieve $\binom{n}{\lfloor n/2 \rfloor}$. On the other hand, since there are 2^n possible sums and only p values, we have by the pigeonhole principle that $\max \Sigma(z) \geq 2^n/p$. To determine $\max \Sigma(z)$ we will use a partition into sparse sets, just as in the proof of theorem 2.3.

Theorem 2.4. *Given $x_1, \dots, x_n \in \mathbb{Z}_p - \{0\}$ for prime p , with $n \leq p-1$, then there exists a symmetric partition of $\mathcal{P}([n])$ into sparse sets with respect to x_1, \dots, x_n .*

Proof. By induction on n . If $\mathcal{C} = \{A_1, \dots, A_k\}$ is sparse with respect to x_1, \dots, x_{n-1} , we want to show that

$$C' = \{A_1, \dots, A_k, A_\ell \cup \{n\}\},$$

$$C'' = \{A_1 \cup \{n\}, A_2 \cup \{n\}, \dots, A_k \cup \{n\}\} - \{A_\ell \cup \{n\}\}$$

are sparse for some ℓ . Set $Y_i = X_{A_i}$ and consider the sums Y_1, \dots, Y_k . We wish to find $1 \leq \ell \leq k$ such that $Y_\ell + x_n \notin \{Y_1, \dots, Y_k\}$. Note

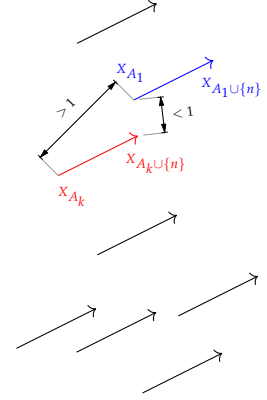


Figure 9: With the choice of A_k shown here, we have that $X_{A_k \cup \{n\}} = X_{A_k} + x_n$ (in red) is near to X_{A_1} (in blue), so C'_i is not a sparse chain, even though C_i is a sparse chain (all the start points of the arrows are far from each other).

[†] So we are choosing A_k to be the furthest in the direction x_n . Then if we add x_n , we are only moving it further away from the others, so the conflict in fig. 9 doesn't occur.

that $k \leq (n-1) + 1 \leq p-1$. Suppose there is no such ℓ . Then wlog, $Y_1 + x_n = Y_2$, $Y_2 + x_n = Y_3, \dots, Y_j + x_n = Y_1$ for some j .

Then $Y_1 + jx_n = Y_1$, so $jx_n = 0$. Then $j = 0$ or $x_n = 0$, which is a contradiction, since $j \leq k \leq p-1$.

In other words, if we have a set $A \subset \Gamma$ for some group Γ , and for some element $x \in \Gamma$ we have $x + A \subset A$. Then A is a union of cosets of a cyclic subgroup of Γ generated by $\{x\}$. \square

Corollary 2.5. *Let $x_1, \dots, x_n \in \mathbb{Z}_p - \{0\}$ with $n \leq p-1$.*

1. *Then $\Sigma(z) \leq \binom{n}{\lfloor n/2 \rfloor}$.*
2. *(Cauchy-Davenport). Let $S(x_1, \dots, x_n) = \{X_A : A \in \mathcal{P}([n])\}$. Then $|S(x_1, \dots, x_n)| \geq \min\{p, n+1\}$.*

Proof.

1. A symmetric partition has exactly $\binom{n}{\lfloor n/2 \rfloor}$ parts, and within each part there is at most one set with corresponding sum z .
2. If $n \leq p-1$, then we have a symmetric partition, which contains a sparse set of size $n+1$. If $n \geq p$, then $S(x_1, \dots, x_n) = \mathbb{Z}_p$. This is because we can just use any $p-1$ elements to get all the elements by sums. \square

References for Section 2.

- Kleitman, D. J. (1970). "On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors". In: *Advances in Mathematics* 5.1, pp. 155–157. ISSN: 0001-8708. DOI: [http://dx.doi.org/10.1016/0001-8708\(70\)90038-1](http://dx.doi.org/10.1016/0001-8708(70)90038-1) (cit. on pp. 3, 14).
- Littlewood, J. E. and A. C. Offord (1938). "On the Number of Real Roots of a Random Algebraic Equation". In: *Journal of the London Mathematical Society* s1-13.4, pp. 288–295. DOI: [10.1112/jlms/s1-13.4.288](https://doi.org/10.1112/jlms/s1-13.4.288) (cit. on pp. 3, 12).
- Erdős, P. (Dec. 1945). "On a lemma of Littlewood and Offord". In: *Bull. Amer. Math. Soc.* 51.12, pp. 898–902. URL: <http://projecteuclid.org/euclid.bams/1183507531> (cit. on p. 12).

3 Intersecting hypergraphs

A family $\mathcal{F} \subset \mathcal{P}([n])$ is *intersecting* if every two sets $A, B \in \mathcal{F}$ have $A \cap B \neq \emptyset$. If \mathcal{F} is intersecting, then $|\mathcal{F}| \leq 2^{n-1}$ because \mathcal{F} can contain at most one set in each pair $\{A, A^c\}$ for every $A \subset [n]$. On the other hand, 2^{n-1} may be achieved by taking $\mathcal{F} = \{A \subset [n] : x \in A\}$ for some $x \in [n]$. What if $\mathcal{F} \subset [n]^{(r)}$ for some r ? If $r > n/2$, then $\mathcal{F} = [n]^{(r)}$ is intersecting.

Theorem 3.1 (Erdős, Ko, and Rado 1961). *Let $r \leq n/2$ and $\mathcal{F} \subset [n]^{(r)}$ be intersecting. Then*

$$|\mathcal{F}| \leq \binom{n-1}{r-1}$$

which can be achieved by $\mathcal{F} = \{A \in [n]^{(r)} : x \in A\}$ for some $x \in [n]$.

Proof. Let us consider a particular circular ordering of $[n]$ and upper bound the number of sets in \mathcal{F} which are intervals of size r in this order. Let us prove there are at most r intervals. Let us fix an interval

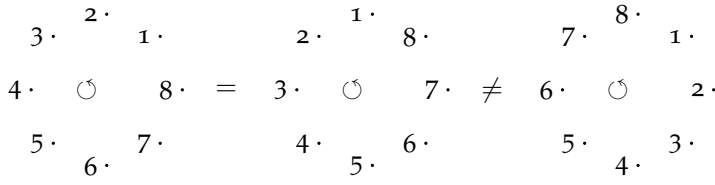


Figure 10: An illustration of circular orders on $[8]$. We define our order counter-clockwise, and so the order is invariant under rotations, as the equality between the left and center orders demonstrates. However, the right order was obtained by reversing the order of the left, and thus is a new order.

$I = (a_1, a_2, \dots, a_r)$, and count the number of intervals which intersect it. For each pair of consecutive points $(a_{i-1}, a_i)^\dagger$ in this interval I , there is an interval I_1 with first element a_{i+1} and an interval I_2 with last element a_i , yielding $2(r-1)$ intervals intersecting it. But \mathcal{F} can contain at most one interval in this pair (I_1, I_2) , because $I_1 \cap I_2 = \emptyset$. So we are left with $r-1$ intervals intersecting I , along with I itself. Thus, we've found r intersecting intervals in this circular order.

How many circular orders are there on $[n]$? Every circular order corresponds to n permutations (all rotations of each other), and there are $n!$ total permutations, yielding $(n-1)!$ circular orders.

In how many circular orders is a given $X \in [n]^{(r)}$ an interval? To obtain an interval, we order the elements of X in $r!$ ways, and then order the rest of the set in $(n-r)!$ ways, yielding $r!(n-r)!$ orders in which X is an interval. See fig. 11 for an illustration.

We have at most r sets of \mathcal{F} as intervals per circular order, and $(n-1)!$ circular orders, so at most $r(n-1)!$ sets in \mathcal{F} over all circular orders.

On the other hand, each set of \mathcal{F} has only $r!(n-r)!$ orders in which it is an interval. Thus, we have $|\mathcal{F}|r!(n-r)!$ intervals corre-

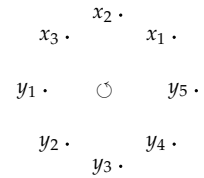


Figure 11: An order in which $X = \{x_1, x_2, x_3\} \in [8]^{(3)}$ is an interval, where we've enumerated $[8] - X = \{y_1, \dots, y_5\}$.

[†] Corresponding to gaps between points

sponding to sets in \mathcal{F} , counted over all orders. Hence,

$$|\mathcal{F}|r!(n-r)! \leq r(n-1)!$$

which completes the proof. \square

Remark. For $r < n/2$, if $\mathcal{F} \subset [n]^{(r)}$ is intersecting and has maximal size, i.e. $|\mathcal{F}| = \binom{n-1}{r-1}$, then for some $x \in n$, we have

$$\mathcal{F} = \{A \in [n]^{(r)} : x \in A\}.$$

The proof is left as an exercise. In particular, there are only n possible extremal families.

Consider $r = n/2$. For each set $A \in [n]^{(r)}$, there is only one forbidden set A^c . So \mathcal{F} may be formed by choosing one set A from each pair $\{A, A^c\}$ arbitrarily. This yields a doubly-exponential number of extremal families.

WE WILL NOW CONSIDER a generalization of intersecting. We say $\mathcal{F} \subset [n]^{(r)}$ is t -intersecting if $|A \cap B| \geq t$ for all $A, B \in \mathcal{F}$. We wish to find $m(n, r, t)$, the maximum size of a t -intersecting family \mathcal{F} which is a subset of $[n]^{(r)}$. A natural guess of an optimal family would be

$$\mathcal{F}_0 := \mathcal{F}_0(n, r, t) = \{A \in [n]^{(r)} : L \subset A\}$$

for some $L \subset [n]$ with $|L| = t$. Note $|\mathcal{F}_0| = \binom{n-t}{r-t}$. On the other hand, for which n, r, t is $m(n, r, t) = \binom{n}{r}$? I.e., when is the family $[n]^{(r)}$ itself t -intersecting? We may use the inequality

$$t \leq |A \cap B| \leq |A| + |B| - |A \cup B| = 2r - n$$

to obtain the bound $2r - t \geq n$. So, as with the 1-intersecting case, for large enough r (compared to t and n), $[n]^{(r)}$ itself is intersecting.

What about $n = 2r - t + 1$? Then $|\mathcal{F}_0| = \binom{2r-2t+1}{r-t}$. But we could still take all r element subsets of $[n-1]$ to obtain

$$m(n, r, t) \geq \binom{2r-t}{r}.$$

But

$$\binom{2r-t}{r} > \binom{2r-t+1}{r} \geq \binom{2r-2t+1}{r-t}.$$

So \mathcal{F}_0 does not achieve the maximum in this case either.

Exercise. Prove that \mathcal{F}_0 is optimal when n is large compared to r and t .

This is the content of theorem 3.2.

Now, consider the family

$$\mathcal{F}_{r-t} := \{A \in [n]^{(r)} : |L \cap A| \geq r\}$$

for some L with $|L| = 2r - t$. Define an interpolation between \mathcal{F}_0 and \mathcal{F}_{r-t} by

$$\mathcal{F}_k := \{A \in [n]^{(r)} : |A \cap L| \geq k + t\}$$

for $|L| = 2k + t$. Then for $A, B \in \mathcal{F}_k$ we have

$$|A \cap B \cap L| \geq |A \cap L| + |B \cap L| - |L| \geq 2(k + t) - (2k + t) = t$$

so for every $0 \leq k \leq r - t$, the family \mathcal{F}_k is t -intersecting.

Example. Consider $r = 5, t = 3$. Then

$$\begin{aligned} |\mathcal{F}_0(n)| &= \binom{n-3}{2}, \\ |\mathcal{F}_1(n)| &= \binom{5}{4} \binom{n-5}{1} + 1 = 5(n-5) + 1, \\ |\mathcal{F}_2(n)| &= \binom{7}{5} = 21. \end{aligned}$$

By choosing different values of n (see table 1), we see that the maximum size family occurs at different k 's: it is a more complicated situation than the 1-intersecting case.

In general, for $n \geq 2k + t$,

$$|\mathcal{F}_k(n)| = \sum_{s=k+t}^{2k+t} \binom{2k+t}{s} \binom{n-2k-t}{r-s}$$

where we are summing over possible sizes s of $|A \cap L|$. The first combination comes from choosing within L , and the second from choosing outside of L .

We may think of $|\mathcal{F}_k(n)|$ as a function of n . Then it is a polynomial of degree $r - k - t$ in which the coefficient of the monomial of largest degree is positive. Among the $\mathcal{F}_k(n)$'s, the family $\mathcal{F}_0(n)$ is eventually of the largest size, because it is the polynomial of highest degree. In fact, $\mathcal{F}_0(n)$ is eventually the largest family overall, as the following result shows.

Theorem 3.2. *For all r, t there exists n_0 such that for $n \geq n_0$,*

$$m(n, r, t) = |\mathcal{F}_0| = \binom{n-t}{r-t}.$$

Proof. Let $\mathcal{F} \subset [n]^{(r)}$ be t -intersecting such that $|\mathcal{F}| = m(n, r, t)$. Then there exists $A, B \in \mathcal{F}$ such that

$$|A \cap B| = t$$

for $n \geq 2r - t$. Assume not. If $\min_{A, B \in \mathcal{F}} |A \cap B| = t + \ell$ for some $\ell \geq 1$, then choose some $L \subset A \cap B$ with $|L| = \ell$, and consider the set

n	k		
	0	1	2
8	10	16	21
11	28	31	21
13	45	41	21

Table 1: The size of $\mathcal{F}_k(n)$, given $r = 5$ and $t = 3$, tabulated over several choices of k and n . We see that the maximal $\mathcal{F}_k(n)$ changes based on both k and n .

$A' = A - L$. If $A' \in \mathcal{F}$, then we would have $|A' \cap B| = t < t + \ell$, so $A' \notin \mathcal{F}$. But for any $C \in \mathcal{F}$, we have

$$t + \ell \leq |A \cap C| = |A' \cap C| + |L \cap C| \leq |A' \cap C| + \ell,$$

so $|A' \cap C| \geq t$. Thus, $\mathcal{F} \cup \{A'\}$ is t -intersecting and $|\mathcal{F} \cup \{A'\}| > |\mathcal{F}|$, which contradicts the maximality of \mathcal{F} .

Now, let $Z = A \cap B$, where $|A \cap B| = t$. If $Z \subset C$ for all $C \in \mathcal{F}$, then $|\mathcal{F}| \leq \binom{n-t}{r-t}$. So we may assume there exists $C \in \mathcal{F}$ such that $Z \not\subset C$. We will show

$$|X \cap (A \cup B \cup C)| \geq t + 1 \quad (\star)$$

for all $X \in \mathcal{F}$. Since each set is of size r , the size $L := |A \cup B \cup C| \leq 3r$. This is enough as it implies

$$|\mathcal{F}| \leq \sum_{s=t+1}^r \binom{L}{s} \binom{n-L}{r-s}$$

which is a polynomial of degree at most $r - t - 1$, and so is eventually less than $\binom{n-t}{r-t}$, which is a polynomial of degree $n - t$.

Let us show (\star) . We know

$$X \cap (A \cup B \cup C) = (X \cap A) \cup (X \cap B) \cup (X \cap C).$$

Each set is of size at least t ; for $|X \cap (A \cup B \cup C)| \leq t$, then $|X \cap (A \cup B \cup C)| = t$, and in particular, $Y := X \cap A = X \cap B = X \cap C$. Then $Y \subset Z$, but since $|Y| = |Z| = t$, we have $Y = Z$. Then we have $Z = Y = X \cap C \subset C$, which is a contradiction. \square

The following theorem, presented here without proof, resolves our question.

Theorem 3.3 (Ahlsweede and Khachatrian 1997). *For all n, r, t ,*

$$m(n, r, t) = \max_k |\mathcal{F}_k|.$$

LET US CONSIDER a consequence of Erdős-Ko-Rado[†]. Let Z_1, \dots, Z_n be independent Bernoulli random variables each with expectation value $p > \frac{1}{2}$. Then $\Pr[Z_i = 1] = p$, $\Pr[Z_i = 0] = 1 - p$. Suppose the Z_i 's are stocks, and the price of each is $\frac{1}{2}$. If we invest \$0.50, then our expected return is $\$p$, no matter how we invest.

Suppose we are very conservative and our goal is to have at least $\frac{1}{2}$ in the end. A good strategy is to diversify and invest uniformly in each stock; then the law of large numbers says that as the number of stocks goes to infinity, we almost surely recover our $1/2$. What's the worst possible strategy? It should be to invest in only 1 stock; in that case, our probability of success is p .

[†] Theorem 3.1

Theorem 3.4 (Liggett 1977). *Let Z_1, \dots, Z_n be independent Bernoulli random variables with expectation value p . Let $c_1, \dots, c_n \geq 0$ with $\sum c_i = 1$. Then*

$$\Pr \left[\sum_{i=1}^n c_i Z_i \geq \frac{1}{2} \right] \geq p.$$

Proof. Assume $c_i > 0$ for all i , and n odd for simplicity. Let $\mathcal{F} = \{A \subset [n] : \sum_{i \in A} c_i \geq \frac{1}{2}\}$. That is, \mathcal{F} is the collection of all sets of r.v. such that if exactly those random variables obtain return 1, we did not lose. Let $\mathcal{F}_k = \mathcal{F} \cap [n]^{(k)}$ and $f_k = |\mathcal{F}_k|$. Then

$$\Pr \left[\sum_{i=1}^n c_i Z_i \geq \frac{1}{2} \right] = \sum_{A \in \mathcal{F}} \Pr[\text{exactly } Z_i \text{ with indices } i \in A \text{ have value } 1].$$

If we fix some A of size k , what is the probability that these k trials succeed? $p^k(1-p)^{n-k}$. Thus,

$$\Pr \left[\sum_{i=1}^n c_i Z_i \geq \frac{1}{2} \right] = \sum_{k=0}^n f_k p^k (1-p)^{n-k}.$$

Our goal is to show that the LHS is larger than p .

Fact 1. \mathcal{F}_k is intersecting for each $k < n/2$.

Proof. Suppose not: then there exists $A, B \in \mathcal{F}_k$ such that $A \cap B = \emptyset$. Then $\sum_{a \in A} c_i \geq \frac{1}{2}$, and $\sum_{a \in B} c_i \geq \frac{1}{2}$. But since each $c_i > 0$, we must have $A \cup B = [n]$, but we know $|A \cup B| < 2 \cdot n/2 = n$. ■

Corollary. For $k < n/2$, we have $f_k \leq \binom{n-1}{k-1}$ by Erdős-Ko-Rado.

This corollary is the essential idea of the proof; the algebraic computation later is long but trivial.

Fact 2. $f_k + f_{n-k} \geq \binom{n}{k}$ for all k .

Proof. For every A , either $A \in \mathcal{F}$ or $[n] - A \in \mathcal{F}$: if the sum of some collection of c_i is less than one half, then the sum of the rest must be at least one half. Thus, for $A \in [n]^{(k)}$, either $A \in \mathcal{F}_k$, or $[n] - A \in \mathcal{F}_{n-k}$, and $f_k + f_{n-k} \geq |[n]^{(k)}| = \binom{n}{k}$. ■

Now,

$$\begin{aligned} \sum_{k=0}^n f_k p^k (1-p)^{n-k} &= \sum_{k < n/2} (f_k + f_{n-k}) p^{n-k} (1-p)^k \\ &\quad + \sum_{k < n/2} f_k (p^k (1-p)^{n-k}) - p^{n-k} (1-p)^k. \end{aligned}$$

By fact 2,

$$\begin{aligned} &\geq \sum_{k < n/2} \binom{n}{k} p^{n-k} (1-p)^k \\ &\quad + \sum_{k < n/2} f_k (p^k (1-p)^{n-k}) - p^{n-k} (1-p)^k. \end{aligned}$$

Since $p^k(1-p)^{n-k} - p^{n-k}(1-p)^k < 0$, the corollary to fact 1 yields

$$\begin{aligned} &\geq \sum_{k < n/2} \binom{n}{k} p^{n-k}(1-p)^k \\ &\quad + \sum_{k < n/2} \binom{n-1}{k-1} (p^k(1-p)^{n-k} - p^{n-k}(1-p)^k). \end{aligned}$$

Now we may group powers of p to obtain

$$\begin{aligned} &= \sum_{k < n/2} p^{n-k} \left[\binom{n}{k} (1-p)^k - \binom{n-1}{k-1} (1-p)^k \right] \\ &\quad + p^k \left[\binom{n-1}{k-1} (1-p)^{n-k} \right]. \end{aligned}$$

We may pull out $(1-p)^k$ of the first term, and use that $\binom{n}{k} - \binom{n-1}{k} = \binom{n-1}{k-1}$ to obtain

$$= \sum_{k < n/2} p^{n-k} \left[\binom{n-1}{k-1} \right] (1-p)^k + p^k \left[\binom{n-1}{k-1} \right] (1-p)^{n-k}.$$

Now, in first term we may switch to summing over $k > n/2$, swapping $n-k$ with k to get

$$\begin{aligned} &= \sum_{k > n/2} p^{n-k} \left[\binom{n-1}{k-1} \right] (1-p)^k \\ &\quad + \sum_{k < n/2} p^k \left[\binom{n-1}{k-1} \right] (1-p)^{n-k}. \\ &= \sum_{k=1}^n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= p \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-1-\ell} \\ &= p(p + (1-p))^{n-1} = p. \end{aligned} \quad \square$$

References for Section 3.

- Ahlsweide, R. and L. H. Khachatrian (Feb. 1997). “The Complete Intersection Theorem for Systems of Finite Sets”. In: *Eur. J. Comb.* 18.2, pp. 125–136. ISSN: 0195-6698. DOI: [10.1006/eujc.1995.0092](https://doi.org/10.1006/eujc.1995.0092) (cit. on p. 20).
- Liggett, T. M. (1977). “Extensions of the Erdős-Ko-Rado theorem and a statistical application”. In: *Journal of Combinatorial Theory, Series A* 23.1, pp. 15–21. ISSN: 0097-3165. DOI: [http://dx.doi.org/10.1016/0097-3165\(77\)90075-9](https://dx.doi.org/10.1016/0097-3165(77)90075-9) (cit. on p. 21).
- Erdős, P., C. Ko, and R. Rado (1961). “Intersection Theorems For Systems Of Finite Sets”. In: *The Quarterly Journal of Mathematics* 12.1, pp. 313–320. DOI: [10.1093/qmath/12.1.313](https://doi.org/10.1093/qmath/12.1.313) (cit. on p. 3, 17).

4 Compression and Shadows

As an aside, let's consider Steiner symmetrization[†], a technique in geometry. Given a shape $K \subset \mathbb{R}^2$, with a line of “symmetry” L , we obtain $S_L(K)$ from K by replacing $K \cap L'$ for every line L' orthogonal to L by an interval of length equal to $|K \cap L'|$ centered at L .

[†] Steiner 1838

Properties:

1. $\text{Area}(S_L(K)) = \text{Area}(K)$
2. $\text{Diam}(S_L(K)) \leq \text{Diam}(K)$
3. $\text{Perimeter}(S_L(K)) \leq \text{Perimeter}(K)$.

The diameter is the largest distance between two points on K . We may prove this by drawing trapezoids between two lines L' and L'' , one of which passes through each point (close to) achieving the diameter.

Given a shape in \mathbb{R}^2 with area 1, what is the smallest diameter? That's the diameter of the area 1 disc. Sketch of proof: take the shape achieving minimum diameter. By a compactness argument, we could show that if we “repeatedly” symmetrize, eventually it is symmetric across every line.

If some bounded shape is symmetric with respect to reflection about three lines L_1, L_2 , and L_3 , then L_1, L_2, L_3 all go through the same point. Why? The center of mass must lie on each line. This will show us that we have a disk. Let us consider a discrete to this symmetrization process: *compression*. For $A \in [n]^{(r)}$, set

$$R_{ij}(A) = \begin{cases} (A - \{j\}) \cup \{i\}, & \text{if } j \in A, i \notin A \\ A, & \text{otherwise.} \end{cases}$$

For example,

$$R_{15}(\{2, 3, 5\}) = \{1, 2, 3\}, \quad R_{15}(\{2, 3, 4\}) = \{2, 3, 4\}, \\ R_{15}(\{1, 3, 5\}) = \{1, 3, 5\}.$$

Let $\tilde{R}_{ij}(\mathcal{A}) = \{R_{ij}(A) : A \in \mathcal{A}\} \cup \{A : R_{ij}(A) \in \mathcal{A}\}$.

Intuition: \tilde{R}_{ij} applies R_{ij} unless the resulting set is already in \mathcal{A} , to prevent collapse.

Properties:

1. $|\tilde{R}_{ij}(\mathcal{A})| = |\mathcal{A}|$.

Let $P_{ij} \subset [n]^{(r)}$ denote the collection of all sets containing j but not i . Then $R_{ij} : P_{ij} \rightarrow P_{ji}$ is bijection.

We will say \mathcal{A} is *compressed* if $\tilde{R}_{ij}(\mathcal{A}) = \mathcal{A}$ for all $i < j$.

2. Any set system \mathcal{A} can be made compressed by applying finitely many compression operators \tilde{R}_{ij} , for $i < j$.

Let $w(A) = \sum_{i \in A} i$, and $w(\mathcal{A}) = \sum_{A \in \mathcal{A}} w(A)$. Then $w(R_{ij}(A)) \leq w(A)$ with equality iff $R_{ij}(A) = A$. Therefore, $w(\tilde{R}_{ij}(\mathcal{A})) \leq w(\mathcal{A})$

with equality iff $\tilde{R}_{ij}(\mathcal{A}) = \mathcal{A}$. Therefore, the process must stop, because we start with finite integer weight and the weight may only decrease (and may not be negative).

In human terms, a compressed set system is one where if we try to replace any element of a set with a smaller element, we end up with something already in the set system. Let us define a natural partial order on $[n]^{(r)}$. If $A = \{a_1, \dots, a_r\}$ with $a_1 < a_2 < \dots < a_r$, and $B = \{b_1, \dots, b_r\}$ with $b_1 < \dots < b_r$, then we say $A \preceq B$ if $a_i \leq b_i$ for all $i = 1, \dots, r$.

3. If $i < j$, then $R_{ij}(A) \preceq A$, and conversely if $A' \preceq A$, then A' can be obtained from A by using finitely many compression operators R_{ij} for $i < j$.

Therefore, \mathcal{A} is compressed if and only if for every $A \prec B$ with $B \in \mathcal{A}$, we have that $A \in \mathcal{A}$. In other words, \mathcal{A} is an ideal in this order. See fig. 12 for an example.

4. If \mathcal{A} is intersecting, then $\tilde{R}_{ij}(\mathcal{A})$ is intersecting. Suppose not. Then there exists $A, B \in \tilde{R}_{ij}(\mathcal{A})$ such that $A \cap B = \emptyset$. If both $A, B \in \mathcal{A}$, then they are intersecting, so let $A \notin \mathcal{A}$. Then $A = R_{ij}(A')$ for some $A' \in \mathcal{A}$ with $j \in A'$. Now, if $B \notin \mathcal{A}$ too, then $B = R_{ij}(B')$ for some $B' \in \mathcal{A}$ with $j \in B'$, and we'd have $i \in A \cap B$, which is a contradiction. So $B \in \mathcal{A}$ with $i \notin B$. If $B = R_{ij}(B)$, then $R_{ij}(B) \in \mathcal{A}$. Otherwise, we have $j \in B$, and we must have $B \neq R_{ij}(B')$ for every $B' \in \mathcal{A}$ (for $B \in P_{ij}$, while the image of R_{ij} is P_{ji} which is disjoint from P_{ij}). In this case then, since $B \in \tilde{R}_{ij}(\mathcal{A})$, we must then have $R_{ij}(B) \in \mathcal{A}$ too. So in either case, $R_{ij}(B) \cap A' \in \mathcal{A}$ which is intersecting, so $R_{ij}(B) \cap A' \neq \emptyset$.

But A' is obtained from A by changing j to i , and $R_{ij}(B)$ is obtained from B by changing j to i , so if $A' \cap R_{ij}(B) \neq \emptyset$, then we have $A \cap B \neq \emptyset$.

Proof of Erdős-Ko-Rado theorem by compression. We wish to show that if $r \leq n/2$, $\mathcal{A} \subset [n]^{(r)}$ is intersecting, then

$$|\mathcal{A}| \leq \binom{n-1}{r-1}.$$

We will proceed by induction on n and r . The case $r = n/2$ is easy: $\binom{n-1}{r-1} = \frac{1}{2} \binom{n}{r}$, and sets in $[n]^{(r)}$ are just complementary pairs.

Assume $r < n/2$. By the facts we have proven, we may assume \mathcal{A} is compressed. Let

$$\mathcal{A}_0 = \{A \in \mathcal{A} : n \notin A\}, \quad \mathcal{A}_1 = \{A \in \mathcal{A} : n \in A\}.$$

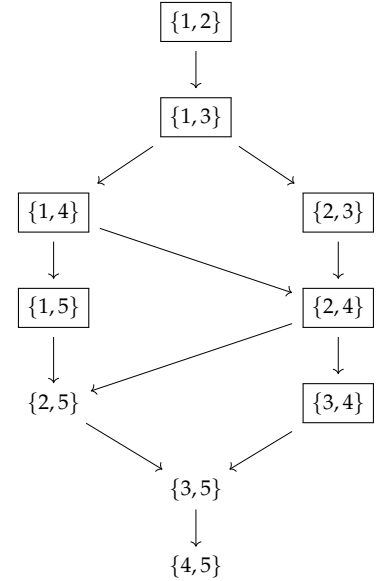


Figure 12: The partial order on $[5]^{(2)}$, where $A \rightarrow B$ means $a \prec b$, and different sets in the same row are incomparable. The boxed elements together form a compressed set.

Then by the IH,

$$|\mathcal{A}| = |\mathcal{A}_0| + |\mathcal{A}_1| \leq \binom{n-2}{r-1} + |\mathcal{A}_1|$$

If $|\mathcal{A}_1| \leq \binom{n-2}{r-2}$, then[†] we have $|\mathcal{A}| \leq \binom{n-1}{r-1}$. Let

[†] using $\binom{n-2}{r-1} + \binom{n-2}{r-2} = \binom{n-1}{r-1}$

$$\mathcal{A}_1 - \{n\} := \{A - \{n\} : A \in \mathcal{A}_1\}.$$

If $\mathcal{A}_1 - \{n\}$ is intersecting, then $|\mathcal{A}_1 - \{n\}| \leq \binom{n-2}{r-2}$ by the IH, since $\mathcal{A}_1 - \{n\} \subset [n-1]^{(r-1)}$. Suppose $\mathcal{A}_1 - \{n\}$ is not intersecting. Then there exists $A, B \in \mathcal{A}$ such that

$$A \cap B = \{n\}.$$

Because $2r < n$, there exists $i < n$ such that $i \notin A \cup B$. Then $R_{in}(B) \in \mathcal{A}$ since \mathcal{A} is compressed. But $A \cap R_{in}(B) = \emptyset$, contradicting that \mathcal{A} is intersecting. \square

Let $\mathcal{A} \subset [n]^{(r)}$, with $|\mathcal{A}| = m$. What is $\min |\partial \mathcal{A}|$? We know the local LYM inequality:

Recall:

$$\partial A = \{B \in [n]^{(r-1)} : B \subset A \text{ for some } A \in \mathcal{A}\}.$$

$$\frac{|\partial A|}{\binom{n}{r-1}} \geq \frac{|\mathcal{A}|}{\binom{n}{r}} \quad (\text{eq. (Local LYM)})$$

but this is rarely tight.

Let's consider $m = 1$, that is, $|\mathcal{A}| = 1$. Then the answer is $r = \binom{r}{r-1}$. For $m = 2$, the answer is $2r - 1$, because their shadows can at most share 1 set. For $r = 2$, we have a simple graph with m edges. What is the minimal number of vertices? We want the minimal n so that $m \leq \binom{n}{2}$. Let us change our question to consider $\mathcal{A} \subset \mathbb{N}^{(r)}$. Given r and m , what is $\min |\partial \mathcal{A}|$ given $\mathcal{A} \subset \mathbb{N}^{(r)}$ with $|\mathcal{A}| = m$?

Lemma 4.1. *We have that*

$$|\partial \tilde{R}_{ij}(\mathcal{A})| \leq |\partial \mathcal{A}|$$

for every $\mathcal{A} \subset \mathbb{N}^{(r)}$.

Proof. It suffices to show that

$$|\partial \tilde{R}_{ij}(\mathcal{A}) - (\partial \mathcal{A})| \leq |\partial \mathcal{A} - (\partial \tilde{R}_{ij}(\mathcal{A}))|.$$

$$\begin{aligned} |X| \geq |Y| &\iff |X - Y| \geq |Y - X| \\ \text{because } |X| &= |X - Y| + |X \cap Y| \text{ and} \\ |Y| &= |Y - X| + |X \cap Y|. \end{aligned}$$

If $B \in \partial \tilde{R}_{ij}(\mathcal{A}) - (\partial \mathcal{A})$, then $i \in B$ but $j \notin B$. Let's see why this is true. Let $A \in \tilde{R}_{ij}(\mathcal{A}) - (\mathcal{A})$ such that $B = A - \{k\}$ for some k . Then $i \in A, j \notin A$ (and so $j \notin B$ too). Now, if $i \notin B$, then $k = i$. But then $B = R_{ji}(A) - \{j\}$ so $B \in \partial \mathcal{A}$, a contradiction.

It is enough to show that

$$R_{ji}(B) \in \partial \mathcal{A} - (\partial \tilde{R}_{ij}(\mathcal{A})).$$

Clearly, $R_{ji}(B) \in \partial\mathcal{A}$, since $B = A - \{k\}$ with $k \neq i, j$, so $R_{ji}(B) = R_{ji}(A) - \{k\}$.

The last piece to check is that $R_{ji}(B) \notin \partial\tilde{R}_{ij}(\mathcal{A})$. If that were the case, then $R_{ji}(A) - \{k\} \subset C$ for some $C \in \tilde{R}_{ij}(\mathcal{A})$. Since $j \in R_{ji}(A)$ and $k \neq j$, we have $j \in C$. Then $R_{ij}(C) \in \mathcal{A}$. But then $B = A - \{k\} \in R_{ij}(C) \in \mathcal{A}$, so $B \in \partial\mathcal{A}$, a contradiction. \square

This lemma shows that for the purposes of minimizing $|\partial\mathcal{A}|$, we may start with a compressed set.

WE DEFINED A PARTIAL ORDER ON $\mathbb{N}^{(r)}$ by $A = \{a_1, \dots, a_r\}$, $a_1 < a_2 < \dots < a_r$, and $B = \{b_1, \dots, b_r\}$, and $b_1 < b_2 < \dots < b_r$, then $A \preceq B$ if $a_i \leq b_i$ for $i \in [r]$. We may define the *lexicographic order* on $\mathbb{N}^{(r)}$. We say $A \leq_L B$ if $(a_1 < b_1)$ or $(a_1 = b_1 \text{ and } a_2 < b_2)$ or $(a_1 = b_1 \text{ and } a_2 = b_2, \text{ but } a_3 < b_3)$ etc. That is, either $A = B$, or if s is the minimal index such that $a_s \neq b_s$, then $a_s < b_s$. See fig. 13 for an example.

Consider the lexicographic order on $\mathbb{N}^{(3)}$. What is the 100th smallest set in this order? $\{1, 2, 102\}$.

Now, let us define the *colexicographic order* on $\mathbb{N}^{(r)}$. Define $A \leq B$ if $a_r < b_r$ or $(a_r = b_r \text{ and } a_{r-1} < b_{r-1})$ or ... etc. We may write this as $A = B$ or if s is the maximal index such that $a_s \neq b_s$, then $a_s < b_s$.

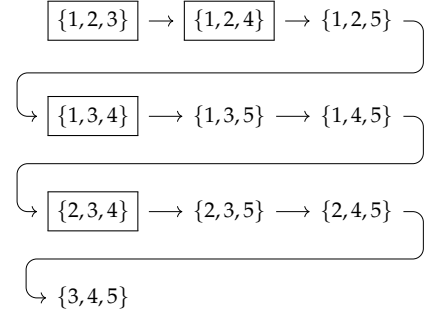


Figure 13: Elements of $[5]^{(3)}$ in lexicographic order, where $A \rightarrow B$ means $A \leq_L B$. The subsets of $[4]^{(3)}$ are boxed; they appear in order, but not next to each other.

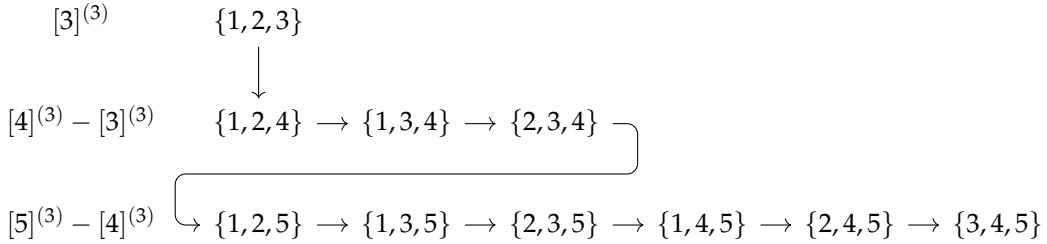


Figure 14: Elements of $[5]^{(3)}$ in colexicographic order. Note that $[n]^{(r)}$ is an initial segment of the colex order on $\mathbb{N}^{(r)}$.

Theorem 4.2 (Katona 1968; Kruskal 1963). *If $\mathcal{A} \subset \mathbb{N}^{(r)}$ with $|\mathcal{A}| = m$, then $|\partial\mathcal{A}|$ is at least as large as the size of the shadow of the first m elements of $\mathbb{N}^{(r)}$ in colexicographic order.*

Example. If $A = \{10, 7, 3\}$ what is the position of A in the colex order on $\mathbb{N}^{(3)}$? In the rows[†] before A , there will be $|[9]^{(3)}| = \binom{9}{3}$ elements. In the row containing A , there are sets of the form $\{10, x, *\}$ with $x < 7$, i.e. $\binom{6}{2}$ sets. Next, there are sets of the form $\{10, 7, y\}$ with $y < 3$, i.e. $\binom{2}{1}$ sets. Finally, we have our set, $\{10, 7, 3\}$. In total then,

$$\binom{9}{3} + \binom{6}{2} + \binom{2}{1} + 1 = 102.$$

See fig. 15 for a diagram.

[†] Referring to an analogous diagram to fig. 14.

$$\{10,6,5\} \rightarrow \{10,7,1\} \rightarrow \{10,7,2\} \rightarrow \{10,7,3\} \rightarrow \{10,7,4\} \rightarrow \{10,7,5\} \rightarrow \{10,7,6\} \rightarrow \{10,8,1\}.$$

For $A \in \mathbb{N}^{(r)}$, let the *initial segment* of A be $I(A) = \{B : B \leq A\}$ and let $i(A) = |I(A)|$.

Lemma 4.3. *If $A = \{a_1, \dots, a_r\}$ with $a_r > a_{r-1} > \dots > a_1$ then*

$$i(A) = \binom{a_r - 1}{r} + \binom{a_{r-1} - 1}{r-1} + \dots + \binom{a_1 - 1}{1} + 1.$$

Proof. $\binom{a_k - 1}{k}$ counts sets $B \subset A$ which coincide with A over

$$a_r, a_{r-1}, \dots, a_{k+1}$$

but whose k th largest element is smaller. \square

Lemma 4.4. *We have that the shadow of the initial segment $I(A)$ is the initial segment of $A - \{\min A\}$. That is,*

$$\partial I(A) = I(A - \{\min A\}).$$

Proof. If $A = \{a_r, \dots, a_2, a_1\}$, then we wish to show that $B \subset \mathbb{N}^{(r-1)}$ is in $\partial I(A)$ if and only if $B \leq \{a_r, \dots, a_2\} = A - \{\min A\}$ in colex order. This is left as an exercise. One way to proceed is by induction on r , splitting into cases where B contains a_r or not. \square

Corollary 4.5. *If*

$$i(A) = \binom{a_r - 1}{r} + \binom{a_{r-1} - 1}{r-1} + \dots + \binom{a_1 - 1}{1} + 1,$$

then

$$|\partial I(A)| = \binom{a_r - 1}{r-1} + \binom{a_{r-1} - 1}{r-2} + \dots + \binom{a_2 - 1}{1} + 1.$$

Let us restate theorem 4.2:

Theorem (Kruskal–Katona). *Let $\mathcal{A} \subset \mathbb{N}^{(r)}$ with*

$$|\mathcal{A}| = \binom{a_r - 1}{r} + \binom{a_{r-1} - 1}{r-1} + \dots + \binom{a_1 - 1}{1} + 1$$

for some $a_r > \dots > a_1$. Then

$$|\partial \mathcal{A}| \geq \binom{a_r - 1}{r-1} + \binom{a_{r-1} - 1}{r-2} + \dots + \binom{a_2 - 1}{1} + 1.$$

Remark. For every positive integer m , we may write m as

$$m = \binom{a_r - 1}{r} + \binom{a_{r-1} - 1}{r-1} + \dots + \binom{a_1 - 1}{1} + 1$$

for some $a_r > a_{r-1} > \dots > a_1$. We have shown this already by considering the m th element in colex order of $\mathbb{N}^{(r)}$.

Figure 15: The colex order near $A = \{10, 7, 3\} \subset \mathbb{N}^{(3)}$. The notation $A \rightarrow B$ means $A \leq B$ in colex order. Since A is the 102nd set in colex order on $\mathbb{N}^{(3)}$, $\{10, 7, 1\}$ is the 100th set which provides a nice comparison to $\{1, 2, 102\}$, the 100th set in lexicographic order. In particular, we note that in colex order, every set is finitely many positions away from the smallest set, while in lexicographic order many sets are infinitely far (like $\{1, 3, 4\}$).

Proof by induction on r ; for fixed r , by induction on $|\mathcal{A}|$. The base case $r = 1$ is trivial. For the induction step, we will assume that \mathcal{A} is compressed, by lemma 4.1. Let

$$\mathcal{A}_1 := \{A \in \mathcal{A} : 1 \in A\}, \quad \mathcal{A}_0 := \mathcal{A} - \mathcal{A}_1.$$

Claim 1: $|\mathcal{A}_1| \geq |\partial \mathcal{A}_0|$.

Compression pushes us towards smaller elements, so \mathcal{A}_1 should be large.

Proof. The map $A \mapsto A \cup \{1\}$ is an injection from $\partial \mathcal{A}_0$ to \mathcal{A}_1 , by compression. ■

Claim 2:

$$|\mathcal{A}_1| \geq \binom{a_r - 2}{r - 1} + \binom{a_{r-1} - 2}{r - 2} + \cdots + \binom{a_2 - 2}{1} + 1.$$

Proof. Assume not. Since $|\mathcal{A}_0| = |\mathcal{A}| - |\mathcal{A}_1|$, we have then have

$$\begin{aligned} |\mathcal{A}_0| &> \left(\binom{a_r - 1}{r} - \binom{a_r - 2}{r - 1} \right) + \left(\binom{a_{r-1} - 1}{r - 1} - \binom{a_{r-1} - 2}{r - 2} \right) + \cdots + \binom{a_1 - 1}{1} \\ &= \binom{a_r - 2}{r} + \binom{a_{r-1} - 2}{r - 1} + \cdots + \binom{a_1 - 2}{1} + 1. \end{aligned}$$

By the IH[†],

[†] Since \mathcal{A} is compressed it must contain 1

$$|\partial \mathcal{A}_0| \geq \binom{a_r - 2}{r - 1} + \cdots + \binom{a_2 - 2}{1} + 1$$

Then claim 1 yields the result. ■

Let $B = \{A - \{1\} : A \in \mathcal{A}_1\} \subset \mathbb{N}^{(r-1)}$. Note $|B| = |\mathcal{A}_1|$.

Claim 3: $|\partial \mathcal{A}| \geq |B| + |\partial B|$.

Proof. Note $B \subset \partial \mathcal{A}$. Let $B' = \{B \cup \{1\} : B \in \partial B\} \subset \partial \mathcal{A}$.

Then $B \cup B' \subset \partial \mathcal{A}$. But since sets in B' contain 1 and sets in B do not contain 1, $|B \cup B'| = |B| + |B'|$. ■

By claims 2, 3, and the IH, we have

$$\begin{aligned} |\partial \mathcal{A}| &\geq |B| + |\partial B| \\ &\geq \binom{a_r - 2}{r - 1} + \binom{a_{r-1} - 2}{r - 2} + \cdots + \binom{a_2 - 2}{1} + 1 \\ &\quad + \binom{a_r - 2}{r - 2} + \binom{a_{r-1} - 2}{r - 3} + \cdots + \binom{a_3 - 2}{1} + 1 \\ &= \binom{a_r - 1}{r - 1} + \cdots + \binom{a_3 - 1}{2} + \binom{a_2 - 1}{1} + 1. \end{aligned} \quad \square$$

Theorem 4.6 (Lovász 1979a). Let $\mathcal{A} \subset \mathbb{N}^{(r)}$, where $|\mathcal{A}| = \binom{x}{r}$ for $x \in \mathbb{R}$. Then $|\partial\mathcal{A}| \geq \binom{x}{r-1}$.[†]

Proof. The proof follows that of the reformulated Kruskal–Katona.

Claims 1 & 3 have the same proof. For claim 2, there is a much simpler proof:

Claim 2. $|\mathcal{A}_1| \geq \binom{x-1}{r-1}$.

Proof. If not, $|\mathcal{A}_0| = |\mathcal{A}| - |\mathcal{A}_1| \geq \binom{x}{r} - \binom{x-1}{r-1} = \binom{x-1}{r}$. Then $|\partial\mathcal{A}_0| \geq \binom{x-1}{r-1} > |\mathcal{A}_1|$, a contradiction. ■

The rest of the proof is the same. □

Corollary 4.7. Let $\mathcal{A} \subset \mathbb{N}^{(r)}$ with $|\mathcal{A}| = \binom{x}{r}$ for $x \in \mathbb{R}$. Let

$$\partial^{(\ell)}\mathcal{A} = \{B \in \mathbb{N}^{(r-\ell)} : B \subset A \text{ for some } A \in \mathcal{A}\}.$$

Then

$$|\partial^{(\ell)}\mathcal{A}| \geq \binom{x}{r-\ell}.$$

Proof by induction on ℓ . The base case is Lovász’s theorem. Then

$$\partial^{(\ell)}\mathcal{A} = \partial(\partial^{(\ell-1)}\mathcal{A})$$

so the IH and Lovász’s theorem yield the result. □

Corollary 4.8. Let G be a 2-graph. Then

$$|G| = |\mathcal{E}(G)| = \binom{x}{2}$$

for some $x \in \mathbb{R}$. Then G contains at most $\binom{x}{k}$ complete subgraphs[‡] of size k .

Remark. We may reformulate this as follows, in a special case. Let G be a graph with $\binom{n}{2}$ edges (but possibly more than n vertices). Then G contains a maximum number of triangles when G is a complete graph of n vertices. We think of this as we are given a budget of edges, and are trying to maximize the number of triangles we make. Here it is intuitive that to do this, we make a complete graph.

Proof. We may assume that $x \geq k$; otherwise we may not form any complete k -subgraphs. If the number of complete subgraphs of G is strictly larger than $\binom{x}{k}$, then it is equal to $\binom{x'}{k}$ for some $x' > x$. So we choose \mathcal{A} to be the family of vertices of complete subgraphs of G :

$$\mathcal{A} = \{V(H) : H \text{ is a complete subgraph of } G\} \subset V(G)^{(k)}.$$

Then each $\{x, y\} \in \partial^{(k-2)}\mathcal{A}$ is a two element subset of a complete subgraph of G , so $\partial^{(k-2)}\mathcal{A} \subset \mathcal{E}(G)$. Then by corollary 4.7, we must have $|\mathcal{E}(G)| \geq \binom{x'}{2} > \binom{x}{2}$, a contradiction. □

We define $\binom{x}{r} := \frac{x(x-1)\cdots(x-r+1)}{r!}$.

Since the polynomials $\binom{x}{r}$ and $\binom{x-1}{r} + \binom{x-1}{r-1}$ agree on the integers, they agree everywhere.

[†] If x is an integer, then equality can be achieved by taking $\mathcal{A} = [x]^{(r)}$.

$\mathcal{E}(G)$ denotes the edge set of G . We identify G with $\mathcal{E}(G)$.

[‡] k -tuples of vertices pairwise joined by edges.

References for Section 4.

- Katona, G. O. H. (1968). "Theory of Graphs: Proceedings of the Colloquium on Graph Theory, Held at Tihany, Hungary, September 1966". In: ed. by P. Erdős and G. O. H. Katona. Academic Press. Chap. A theorem of finite sets (cit. on p. 26).
- Kruskal, J. B. (1963). "Mathematical optimization techniques". In: ed. by R. Bellman. University of California Press. Chap. The Number of Simplices in a Complex, p. 251 (cit. on p. 26).
- Lovász, L. (1979a). *Combinatorial Problems and Exercises*. AMS/Chelsea publication. Problem 13.31. North-Holland Publishing Company. ISBN: 9780821869475 (cit. on p. 29).
- Steiner, J. (1838). "Einfache Beweise der isoperimetrischen Hauptsätze." In: *Journal für die reine und angewandte Mathematik* 18, pp. 281–296 (cit. on p. 23).

5 Turán type problems

The general problem we've been considering is to find $\max |\mathcal{F}|$ given that $\mathcal{F} \subset \mathcal{P}([n])$ has certain properties. For example, Sperner systems, and intersecting r -graphs. In these problems, we are forbidding certain sub set systems[†]. Here, we will focus on r -graphs.

We will say \mathcal{F} and \mathcal{H} are *isomorphic* set systems if there exists a bijective map $\phi : \bigcup_{F \in \mathcal{F}} F \rightarrow \bigcup_{H \in \mathcal{H}} H$ such that $F \in \mathcal{F}$ if and only if $\phi(F) \in \mathcal{H}$. Now, let forbidden subconfigurations F_1, \dots, F_k be given r -graphs[‡]. Define the *Turán number*

$$\text{ex}(n; F_1, \dots, F_k)$$

to be the maximum $|\mathcal{F}|$ such that \mathcal{F} does not contain a subgraph isomorphic to any of F_1, F_2, \dots, F_k .

Remark. We will increasingly refer to $F \in \mathcal{F}$ as edges.

Example. If M_2 consists of two disjoint edges of size r , then Erdős-Ko-Rado says that $\text{ex}(n; M_2) = \binom{n-1}{r-1}$ for $n \geq 2r$.

Let

$$\pi(n; F_1, \dots, F_k) := \frac{\text{ex}(n; F_1, \dots, F_k)}{\binom{n}{r}}$$

be the ratio of the largest valid size of \mathcal{F} to the size of $[n]^{(r)}$. We define the *Turán density* of F_1, \dots, F_k to be

$$\pi(F_1, \dots, F_k) := \lim_{n \rightarrow \infty} \pi(n; F_1, \dots, F_k) \in [0, 1].$$

Example.

$$\pi(M_2) = \lim_{n \rightarrow \infty} \frac{\binom{n-1}{r-1}}{\binom{n}{r}} = 0,$$

using Erdős-Ko-Rado.

Theorem 5.1 (Katona, Nemetz, and Simonovits 1964). *If $r \leq n_0 \leq n$, then*

$$\pi(n_0; F_1, \dots, F_k) \geq \pi(n; F_1, \dots, F_k).$$

Remark. Then $\pi(n; F_1, \dots, F_k)$ decreases for $n \geq r$ and is bounded below by zero, so $\pi(F_1, \dots, F_k)$ exists.

Proof. Let $\mathcal{F} \subset [n]^{(r)}$ be such that $|\mathcal{F}| = \text{ex}(n; F_1, \dots, F_k)$ and \mathcal{F} contains no subgraphs isomorphic to any of F_1, \dots, F_k . Let H_1, H_2, \dots, H_N be the restrictions* of \mathcal{F} to all possible n_0 element subsets of $[n]$, where $N = \binom{n}{n_0}$. It suffices to show that

$$\pi(n; F_1, \dots, F_k) = \frac{|\mathcal{F}|}{\binom{n}{r}} \leq \frac{1}{\binom{n}{n_0}} \sum_{i=1}^N \frac{|H_i|}{\binom{n_0}{r}}$$

[†] In a Sperner system, there is no 2 element sub system $\{A, B\}$ with $A \subset B$. In an intersecting r -graph, there is no two element sub system $\{A, B\}$ with $A \cap B = \emptyset$.

[‡] Recall: Set systems with elements which are in $[n]^{(r)}$.

Not having a subgraph isomorphic to M_2 means there must not be two disjoint subsets in \mathcal{F} , i.e., \mathcal{F} is intersecting.

If you select two r -tuples of elements in an n element set for very large n , you almost surely get disjoint tuples.

* Note that for $X \subset [n]$, we define the restriction of \mathcal{F} as $\mathcal{F}|_X := \{A \in \mathcal{F} : A \subset X\}$.

since $\frac{|H_i|}{\binom{n_0}{r}} \leq \pi(n_0; F_1, \dots, F_k)$.

We are left to estimate $\sum_{i=1}^N |H_i|$. Given an edge in $F \in \mathcal{F}$, how many H_i 's does it appear in? Each H_i whose base set[†] includes F , so we need to choose $n_0 - r$ more elements for the base set from the $n - r$ remaining possible elements, i.e. $\binom{n-r}{n_0-r}$. Then,

$$\frac{1}{\binom{n}{n_0}} \sum_{i=1}^N \frac{|H_i|}{\binom{n_0}{r}} = |\mathcal{F}| \frac{\binom{n-r}{n_0-r}}{\binom{n}{n_0} \binom{n_0}{r}} \equiv \frac{|\mathcal{F}|}{\binom{n}{r}}.$$

To show the boxed equality and finish the proof, we need

$$\binom{n}{r} \binom{n-r}{n_0-r} = \binom{n}{n_0} \binom{n_0}{r}. \quad (1)$$

This is an identity which we can combinatorially reason as follows. The RHS means we first choose n_0 elements out of n , then choose r out of those n_0 . The LHS means we choose r elements from n then $n_0 - r$ elements from the remaining $n - r$. This is depicted in fig. 16. \square

Let H be a graph (2-graph). Then $\text{ex}(n, H) = \max |\mathcal{E}(G)|$, where the maximum is taken over graphs G with $|V(G)| = n$ such that H is not a subgraph of G .

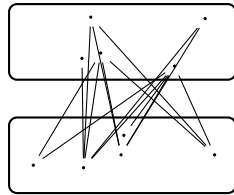
We may consider

$$\pi(n; H) := \frac{\text{ex}(n, H)}{\binom{n}{2}}, \quad \pi(H) := \lim_{n \rightarrow \infty} \pi(n, H).$$

By theorem 5.1, $\pi(n, H)$ decreases for $n \geq 2$ (and fixed H), so $\pi(H)$ exists. Let K_n be the complete graph on n vertices: $K_n = [n]^{(2)}$. Since K_2 is just a single edge, $\text{ex}(n; K_2) = 0$.

Consider $P_3 = \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \end{array}$. Then $\text{ex}(n, P_3) = \lfloor n/2 \rfloor$, and $\pi(P_3) = 0$.

Consider $K_3 = \triangle$. Let us bound $\pi(K_3)$. Consider n even, and the complete bipartite graph on n vertices, shown in fig. 17.



Then $\pi(n, K_3) \geq \frac{n^2/4}{\binom{n}{2}} \rightarrow \frac{1}{2}$. On the other hand, P_3 achieves

$$\pi(3, K_3) = \frac{\text{ex}(3, K_3)}{3} = \frac{2}{3},$$

This estimate is the only actual inequality in the proof.

[†] Meaning the n_0 -element set X such that $H_i = \mathcal{F} \upharpoonright X$.

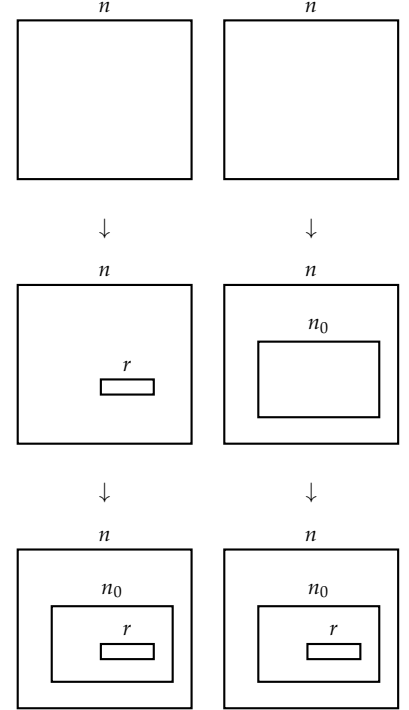


Figure 16: (above) *Left*: an illustration of the LHS of eq. (1) and *right*: the RHS of eq. (1).

Figure 17: (left) The complete bipartite graph on n vertices. We divide the graph into independent sets of size $n/2$, then connect each vertex in the upper set to each vertex in the lower set. This produces $(n/2)^2$ edges and no triangles.

so by theorem 5.1, $\pi(K_3) \leq \frac{2}{3}$. Thus, we have

$$\frac{1}{2} \leq \pi(K_3) \leq \frac{2}{3}.$$

How do we make a graph with many edges that does not contain any complete subgraphs on t vertices?

We look at t groups of size $\frac{n}{t}$. Then join two vertices if they lie in different groups, but not join them if they lie in the same group.

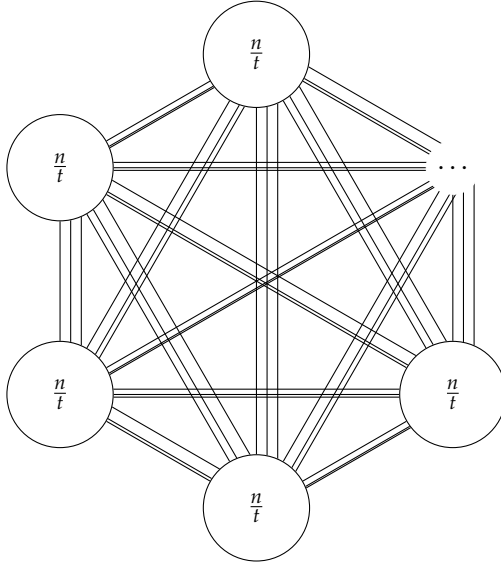


Figure 18: The Turán graph on n vertices, in the case that n is divisible by t . We partition the graph into t independent subsets of size $\frac{n}{t}$. Then we connect each vertex in each independent subset A to all the vertices in $V(G) - A$.

The *Turán graph* $T_t(n) = T$ is a graph with $|V(T)| = n$ such that $V(T)$ is partitioned into A_1, \dots, A_t and $v \in A_i$ is adjacent to $u \in A_j$ iff $i \neq j$, and the sizes obey $||A_i| - |A_j|| \leq 1$ for all i, j .

Then

$$\pi(K_t) \geq \lim_{n \rightarrow \infty} \frac{|\mathcal{E}(T_{t-1}(n))|}{\binom{n}{2}} = \lim_{n \rightarrow \infty} \frac{\binom{n}{2} - (t-1)\binom{\frac{n}{t-1}}{2}}{\binom{n}{2}} = 1 - \frac{1}{t-1} = \frac{t-2}{t-1}$$

Theorem 5.2 (Turán 1941). For every $t \geq 2$, $\pi(K_t) = \frac{t-2}{t-1}$. Moreover,

$$\text{ex}(n, K_t) \leq \frac{t-2}{t-1} \frac{n^2}{2}.$$

Proof by induction on n . Note that when n is divisible by $t-1$,

$$\text{ex}(n, K_t) \geq |\mathcal{E}(T_{t-1}(n))| = \frac{t-2}{t-1} \frac{n^2}{2}$$

and equality is achieved. Base case: for $n < t-1$, then the maximal number of edges (without restriction) is $\binom{n}{2} = \frac{n-1}{n} \frac{n^2}{2} \leq \frac{t-2}{t-1} \frac{n^2}{2}$.

Induction step. Let $n \geq t-1$. We may assume that K_{t-1} is a subgraph of our graph G (where G is a graph on vertices with no K_t subgraph and $|\mathcal{G}| = \text{ex}(n, K_t)$).

Let $U = \{v_1, v_2, \dots, v_{t-1}\}$ be the set of vertices of this K_{t-1} subgraph. Then

$$|\mathcal{E}(G)| = \frac{(t-1)(t-2)}{2} + |\mathcal{E}(U, V(G) - U)| + |\mathcal{E}(G - U)|.$$

This is the number of edges within U , plus the number of edges with exactly one end in U , plus the number of edges not connected to U , respectively. See fig. 19 for a depiction of this partition.

For every $u \in V(G) - U$, the vertex u is adjacent to $\leq t-2$ vertices in U (otherwise $U \cup \{u\} \cong K_t$). So

$$|\mathcal{E}(U, V(G) - U)| \leq (t-2)|V(G) - U| = (t-2)(n-t-1).$$

Moreover,

$$|\mathcal{E}(G - U)| \leq \frac{t-2}{t-1} \frac{(n-t+1)^2}{2},$$

by the induction hypothesis. Putting it all together,

$$\begin{aligned} |\mathcal{E}(G)| &\leq \frac{(t-1)(t-2)}{2} + (t-2)(n-t+1) + \frac{t-2}{t-1} \frac{(n-t+1)^2}{2} \\ &= \frac{t-2}{2(t-1)} \left((t-1)^2 + 2(t-1)(n-t+1) + (n-t+1)^2 \right) \\ &= \frac{t-2}{2(t-1)} n^2. \end{aligned} \quad \square$$

We'll provide another proof of Turán's theorem, but first let us introduce some notation. Let

$$d(G) = \frac{2|\mathcal{E}(G)|}{n^2}$$

be the *density* of G ; this is the probability that choosing two vertices uniformly at random (with repetition) from $V(G)$ gives an edge. The density λ is called the *Lagrangian* of G .

Suppose $V(G) = [n]$. Let

$$\lambda(G) := \max_{\substack{x_i \geq 0, \\ \sum_{i=1}^n x_i = 1}} \sum_{(i,j) \in \mathcal{E}(G)} x_i x_j.$$

Then $2\lambda(G)$ is the maximum probability of selecting an edge by independently sampling two vertices taken over all probability distributions on the vertex set. In particular, $2\lambda(G) \geq d(G)$, for every G .

Example.

$$\lambda(K_2) = \max_{\substack{x_1, x_2 \geq 0, \\ x_1 + x_2 = 1}} x_1 x_2 = \max_{x_1 \geq 0} x_1(1 - x_1) = \frac{1}{4},$$

achieved when $x_1 = x_2 = \frac{1}{2}$.

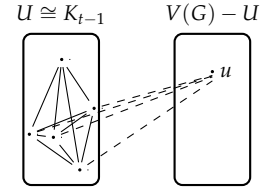


Figure 19: Illustration of the partition of G into $V(G) - U \ni u$ and U . If u were adjacent to more than $t-2$ notes in U , then $U \cup \{u\} \cong K_t$.

$$\lambda(P_3) = \max_{\substack{x_1, x_2, x_3 \geq 0 \\ x_1 + x_2 + x_3}} x_1 x_2 + x_2 x_3 = \max_{\substack{x_1, x_2, x_3 \geq 0 \\ x_1 + x_2 + x_3}} x_2(x_1 + x_3)$$

is still a product of two things which sum to one. So we need $x_2 = \frac{1}{2}$ and $x_1 + x_3 = \frac{1}{2}$. We could take $x_1 = x_3 = \frac{1}{4}$.

Lemma 5.3. $\lambda(K_t) = \frac{t-1}{2t}$.

Proof. The uniform distribution on $|V(K_t)|$ achieves $\frac{t-1}{2t}$, so it is enough to show

$$\sum_{\substack{1 \leq i < j \leq t, \\ x_i \geq 0, \\ \sum_{i=1}^n x_i = 1}} x_i x_j \leq \frac{t-1}{2t}.$$

But

$$\sum_{\substack{1 \leq i < j \leq t, \\ x_i \geq 0, \\ \sum_{i=1}^n x_i = 1}} 2x_i x_j = (x_1 + x_2 + \cdots + x_t)^2 - \sum_{i=1}^t x_i^2 = 1 - \sum_{i=1}^t x_i^2 \quad \boxed{\leq} \quad 1 - \frac{1}{t}$$

where we need to show the boxed inequality. Equivalently, we need $\sum_{i=1}^t x_i^2 \geq \frac{1}{t}$ for all x_i as above. But this follows from Jensen's inequality (with the uniform distribution): if f is convex, then

$$\frac{\sum_{i=1}^n f(x_i)}{n} \geq f\left(\frac{x_1 + \cdots + x_n}{n}\right)$$

for all x_1, \dots, x_n .

Here, we take $f(x) = x^2$, to obtain

$$\frac{\sum_{i=1}^n x_i^2}{t} \geq \left(\frac{x_1 + \cdots + x_t}{t}\right)^2 = \left(\frac{1}{t}\right)^2. \quad \square$$

Theorem 5.4. If G has no K_t subgraph, then $\lambda(G) \leq \lambda(K_{t-1}) = \frac{t-2}{2(t-1)}$.

Proof. Let $p_G(\bar{x}) = \sum_{\{i,j\} \in \mathcal{E}(G)} x_i x_j$. Then

$$\lambda(G) = \max_{\substack{x_i \geq 0 \\ \sum_i x_i = 1}} p_G(\bar{x}).$$

Choose maximal \bar{x} so that $p_G(\bar{x}) = \lambda(G)$, and $\#\{i : x_i \neq 0\}$ is minimal. We may assume that in fact $x_i > 0$ for all i , by throwing away vertices with zero weights.

Claim. G is complete.

Remark. This means the probability distribution was concentrated on a complete subgraph.

Proof of claim. Suppose G is not complete. Then there exists $i, j \in V(G)$ non-adjacent. We have

$$p_G(\bar{x}) = x_i \overbrace{\sum_{\substack{k: \\ \{k,i\} \in \mathcal{G}(G)}}^{C_i}} x_k + x_j \overbrace{\sum_{\substack{k: \\ \{k,j\} \in \mathcal{G}(G)}}^{C_j}} x_k + \overbrace{\sum_{\substack{\{k,\ell\} \in \mathcal{E}(G): \\ k,\ell \in V(G) - \{i,j\}}}^b} x_k x_\ell.$$

Assume wlog that $C_j \geq C_i$. Let \bar{x}' be obtained by setting $x'_i = 0$ and $x'_j + j = x_i + x_j$, and the other $x'_k = x_k$ (for $k \neq j$ and $k \neq i$). Then $p_G(\bar{x}') = (x_i + x_j)C_j + b \geq x_i C_i + x_j C_j + b = p_G(\bar{x})$.

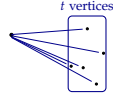
This is a contradiction: \bar{x}' has more zero values than \bar{x} , but still achieves $\lambda(G)$. But we choose \bar{x} to have the minimal number of non-zero values. ■

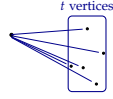
Then G is complete, and so must be of size $t - 1$. □

Remark. Theorem 5.4 proves theorem 5.2.

Proof.


$$\frac{|\mathcal{E}(G)|}{n^2} = p_G\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \leq \lambda(G) \leq \frac{t-2}{2(t-1)}. \quad \square$$



Let $K_{1,t}$ be the graph . Then $\pi(K_{1,t}) = 0$, as follows: We may bound

$$\text{ex}(n, K_{1,t}) \leq \frac{(t-1)n}{2}$$

because every vertex has degree $\leq t - 1$ if there is no $K_{1,t}$ subgraph.

Let $K_{2,t}$ be the graph . If G has no $K_{2,t}$ then it has $\leq (t-1)\binom{n}{2}$ paths P_3 as subgraphs, but if G has $\epsilon\binom{n}{2}$ edges, we “expect” $\geq \epsilon^2\binom{n}{3}$ paths P_3 , so if $\epsilon > 0$ for large n , we get a contradiction. Thus, $\pi(K_{2,t}) = 0$.

Theorem 5.5. $\pi(K_{t,t}) = 0$ for every $t > 1$.

Remark. This implies $\pi(H) = 0$ for every bipartite graph H .

Proof. We need to show that for every $\epsilon > 0$ there exists n_0 such that if G has no $K_{t,t}$ subgraph, and $n \geq n_0$ vertices, then $|\mathcal{E}(G)| \leq \epsilon\binom{n}{2}$.

Suppose that $|\mathcal{E}(G)| \geq \epsilon\binom{n}{2}$.

Set

$$f(G) = \sum_{\{v_1, v_2, \dots, v_t\} \subset V(G)^{(t)}} |N(v_1) \cap N(v_2) \cdots \cap N(v_t)| \leq (t-1) \binom{n}{t}$$

where $N(v)$ is the set of neighbors of v in G : that is, $N(v) = \{u \in V(G) : \{u, v\} \in \mathcal{E}(G)\}$.

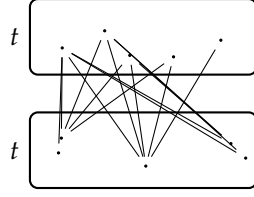
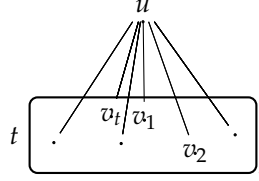


Figure 20: Left. The bipartite graph on n vertices. We divide the graph into independent sets of size $n/2$, then connect each vertex in the upper set to each vertex in the lower set. This produces $(n/2)^2$ edges and no triangles.



We are counting subgraphs of the form

First, note

$$\binom{m}{t} \geq \frac{m^t}{t!} - cm^{t-1}$$

for some constant c depending on t only.

Then,

$$f(G) = \sum_{u \in V(G)} \binom{\deg(u)}{t} \geq \sum_{u \in V(G)} \left(\frac{\deg^t(u)}{t!} - c \deg^{t-1}(u) \right)$$

where we've used Jensen's for t th powers. Then,

$$\begin{aligned} &\geq \frac{n}{t!} \left(\sum_{u \in V(G)} \frac{\deg(u)}{n} \right)^t - cn^t \\ &\geq \frac{n}{t!} \left(\frac{2\epsilon \binom{n}{2}}{n} \right)^t - cn^t \end{aligned}$$

Using $2\epsilon \binom{n}{2} \geq \frac{\epsilon}{2} n^2$

$$\begin{aligned} &\geq \frac{n}{t!} \left(\frac{\epsilon}{2} n \right)^t - cn^t \\ &\boxed{\geq} (t-1) \binom{n}{t} \end{aligned}$$

where the boxed inequality yields a contradiction, and holds for large enough t . \square

If H is not bipartite, is it possible $\pi(H) = 0$? No, there exist large “dense” graphs with no H subgraph. For every non-bipartite graph, the Turán density is at least $1/2$, for the same reason as K_3 : it cannot be embedded in a large complete bipartite graph, so these graphs[†] witness this.

Let us consider graphs which are not subgraphs of the Turán graph $T_3(n)$ (depicted in fig. 23): If H is not a subgraph of $T_3(n)$ for any n , then $\pi(H) \geq \frac{2}{3}$. Let us generalize.

[†] which have density $1/2$

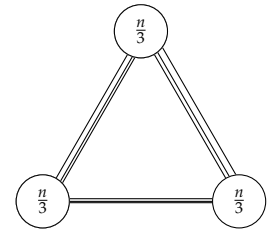


Figure 21: The Turán graph $T_3(n)$.

We say $c: V(G) \rightarrow [k]$ is a k -coloring if $c(u) \neq c(v)$ for every $\{u, v\} \in \mathcal{E}(G)$.

We write $\chi(G)$ for the minimum k such that G admits a k -coloring. With this definition in hand, we may formulate the following result.

Lemma 5.6. *If H contains an edge,*

$$\pi(H) \geq \frac{\chi(H) - 2}{\chi(H) - 1}.$$

Proof. Let $k = \chi(H) - 1$. As H is not k -colorable, H is not a subgraph of any Turán graph $T_k(n)$ and so $\pi(H) \geq \lim_{n \rightarrow \infty} \frac{|\mathcal{E}(T_k(n))|}{\binom{n}{2}} = \frac{k-1}{k}$. \square

Lemma 5.7. *For every r, n_0 , and ϵ , there exists N such that if G is an r -graph with $|V(G)| = n \geq N$ and $|\mathcal{E}(G)| \geq d_r(n)$, then G contains a subgraph (sub r -graph) G' such that*

$$|V(G')| = n' \geq n_0$$

and every vertex $v \in V(G')$ belongs to at least $(d - \epsilon)\binom{n'}{r-1}$ edges.

Proof. Suppose not. Then there exists a vertex $v_1 \in V(G)$ such that v_1 belongs to at most $(d - \epsilon)\binom{n}{r-1}$ edges. Delete this vertex to obtain a graph G_1 which in turn has a vertex v_2 in at most $(d - \epsilon)\binom{n-1}{r-1}$ edges. Delete this vertex to obtain G_2 , and continue in the same manner.

We eventually arrive at a graph G_{n-n_0} on n_0 vertices. Then

$$\begin{aligned} d\binom{n}{r} &\leq |G| \\ &\leq (d - \epsilon)\binom{n}{r-1} + (d - \epsilon)\binom{n-1}{r-1} + \cdots + (d - \epsilon)\binom{n_0+1}{r-1} + \underbrace{|G_{n-n_0}|}_{\leq \binom{n_0}{r}}. \end{aligned}$$

It remains to show that for n large enough (in terms of n_0, r, ϵ)

$$d\binom{n}{r} > (d - \epsilon) \left[\binom{n}{r-1} + \binom{n-1}{r-1} + \cdots + \binom{n_0+1}{r-1} \right] + \binom{n_0}{r}.$$

But, repeating the Pascal's triangle inequality,

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-2}{r-1} + \binom{n-3}{r-1} + \cdots + \binom{r-1}{r-1}.$$

So,

$$d\binom{n}{r} = (d - \epsilon) \left(\binom{n-1}{r-1} + \binom{n-2}{r-1} + \binom{n-3}{r-1} + \cdots + \binom{r-1}{r-1} \right) + \epsilon\binom{n}{r}.$$

This eliminates most terms; we are left with

$$\epsilon\binom{n}{r} \stackrel{?}{>} (d - \epsilon)\binom{n}{r-1} + \binom{n_0}{r}.$$

So we can restrict to a (large) subgraph to obtain a minimal bound on vertex degree, at the cost of ϵ density.

But the polynomial in n on the left has degree r , and on the right, degree $r - 1$, so for $n \geq N$ with N large enough, we have strict inequality.

□

Theorem 5.8 (Erdős and Stone 1946).

$$\pi(H) = \frac{\chi(H) - 2}{\chi(H) - 1}$$

for every 2-graph H with $\chi(H) \geq 2$.

I.e. H contains an edge.

Proof. Lemma 5.6 gives the lower bound. Now, consider $K_{\underbrace{t, \dots, t}_{k \text{ times}}}$ the complete k -partite graph with parts of size t , as depicted in fig. 22.

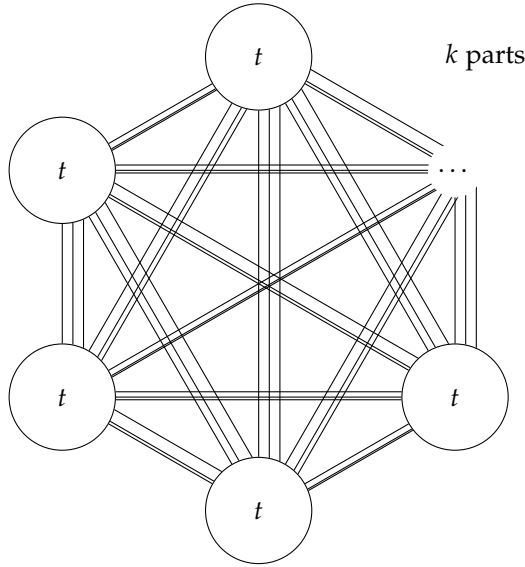


Figure 22: An illustration of $K_{t, \dots, t}$, where there are k t 's in the subscript. This means we have k independent sets of size t , and each vertex of each independent set is connected to all the vertices in all the other sets.

For every t ,

$$\pi(K_{t, \dots, t}) \leq \frac{k-2}{k-1}$$

by induction on k . Suppose for some k , and some t ,

$$\pi(K_{\underbrace{t, \dots, t}_{k \text{ times}}}) \geq \frac{k-2}{k-1} + \epsilon.$$

for $\epsilon > 0$. It is enough to show that for every graph G with n vertices such that

$$|\mathcal{E}(G)| \geq \left(\frac{k-2}{k-1} + \epsilon \right) \binom{n}{2},$$

we must have that G contains $K_{\underbrace{t, \dots, t}_{k \text{ times}}}$.

By lemma 5.7 we may assume that every vertex of G has degree

$$\geq \left(\frac{k-2}{k-1} + \frac{\epsilon}{2} \right) n.$$

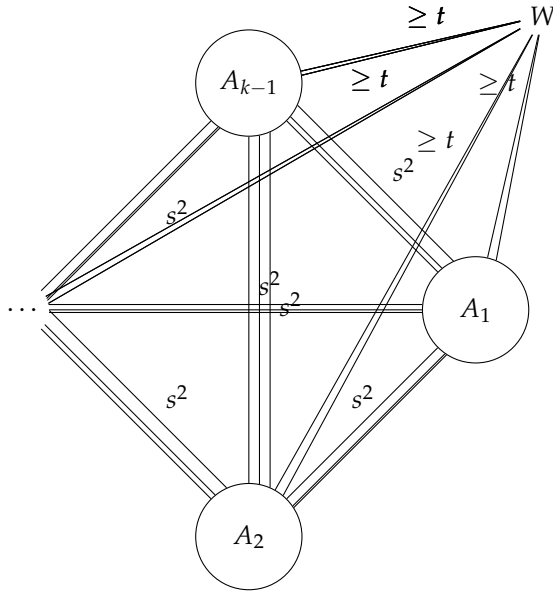
By the induction hypothesis, G contains $\underbrace{K_{s,\dots,s}}_{k-1 \text{ times}}$ for every s . We

will choose a particular s depending only on k and ϵ , which we will specify later.

It suffices to prove that if n is large compared to s, k, ϵ , and s is large compared to k, t and ϵ , and G is a graph with n vertices, and each vertex has degree at least $\left(\frac{k-2}{k-1} + \frac{\epsilon}{2} \right) n$, and G contains a complete $k-1$ -partite subgraph with s vertices in each part, then G contains a complete k -partite subgraph with t vertices.

Let $A_1, A_2, \dots, A_{k-1} \subset V(G)$ with $|A_i| = s$ such that every vertex of A_i is adjacent to every vertex of A_j when $i \neq j$. Let $U = A_1 \cup \dots \cup A_{k-1}$ and let W be the set of vertices in $V(G)$ which have $\geq t$ neighbors in each A_i .

Such a group of sets is simply $\underbrace{K_{s,\dots,s}}_{k-1 \text{ times}}$ and exists by the induction hypothesis.



Now, we wish to show $w := |W|$ is large. First,

$$\begin{aligned} \left(\frac{k-2}{k-1} + \epsilon \right) n(k-1)s &\leq \sum_{v \in U} \deg(v) \leq \sum_{v \in V(G)} |N(v) \cap U| \\ &\leq \underbrace{ws(k-1)}_{\text{from vertices in } W} + \underbrace{(n-w)(s(k-2)+t)}_{\text{from vertices not in } W} \\ ((k-2) + \epsilon(k-1))ns &\leq ns(k-2) + nt + w(s-t) \end{aligned}$$

If we choose s such that $\epsilon(k-1)s - t \geq 1$, then

$$n \leq n(\epsilon(k-1)s - t) \leq w(s-t) \leq ws.$$

Then $|W| \geq \frac{n}{s}$. For each $v \in W$ let $(B_1^v, B_2^v, \dots, B_{k-1}^v)$ be the sets of neighbors of v of size t such that $B_i \subset A_i$. There are $\binom{s}{t}^{k-1}$ choices of these sequences of neighbors, so if $w > \underbrace{s(k-1)}_{\text{vertices in } U} + (t-1)\binom{s}{t}^{k-1}$ then So $|W - U| \geq |W| - |U| \geq (t-1)\binom{s}{t}^{k-1}$.

by pigeonhole principle, there exists t entries in $W - U$ which have the same t neighbors in each of the A_i , as desired. \square

The entries in $W - U$ don't need to be independent, because we just need a subgraph, so we can just not include those edges in our subgraph.

References for Section 5.

- Katona, G. O. H., T. Nemetz, and M. Simonovits (1964). "On a graph problem of Turán." Hungarian. In: *Mat. Lapok*, pp. 228–238 (cit. on p. 31).
- Erdős, P. and A. H. Stone (Dec. 1946). "On the structure of linear graphs". In: *Bull. Amer. Math. Soc.* 52.12, pp. 1087–1091. DOI: [10.1090/S0002-9904-1946-08715-7](https://doi.org/10.1090/S0002-9904-1946-08715-7) (cit. on p. 39).
- Turán, P. (1941). "On an extremal problem in graph theory". In: *Mat. Fiz. Lapok* 48.436-452, p. 137 (cit. on p. 33).

6 Hypergraph Turán Problems

Let $K_4^{(3)}$ be a complete 3-graph on 4 vertices:

$$K_4^{(3)} = \{\{a, b, c\}, \{a, b, d\}, \{a, c, d\}, \{b, c, d\}\}.$$

Then $\pi(K_4^{(3)})$ is not known[†].

Consider the generalized triangle $T_3 = \{D_1, D_2, D_3\}$ where $D_1 = \{1, 2, 3\}$, $D_2 = \{1, 2, 4\}$, and $D_3 = \{3, 4, 5\}$; see fig. 23.

We will say $C : V(H) \rightarrow [k]$ is a *strong k -coloring* if $c(v) \neq c(u)$ whenever u and v belong to an edge of H . $\chi_s(H) = \min k$ such that H can be strongly k -colored.

In this language, $\chi_s(T_3) = 4$, so T_3 is not a subgraph of any strongly 3-colorable graph. There exist strongly 3-colorable graphs with density $2/9$, so $\pi(T_3) \geq 2/9$. We will show $\pi(T_3) = \frac{2}{9}$.

First, we will define an analog to the 2-graph Lagrangian defined in the previous section. Let H be an r -graph, $V(H) = [n]$. For $\bar{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, let

$$P_H(\bar{x}) = \sum_{e \in H} \prod_{i \in e} x_i.$$

If $x_i \geq 0$ and $\sum_i x_i = 1$, i.e. \bar{x} corresponds to a probability distribution on $|V(H)|$, then $\frac{1}{r!} P_H(\bar{x})$ is the probability that selecting r vertices independently at random produces an edge. The *hypergraph Lagrangian* is $\lambda(H) := \max P_H(\bar{x})$ where the maximum is taken over all $\bar{x} \geq 0$, $\sum_{i \in [n]} x_i = 1$.

Lemma 6.1. *Let H be an r -graph with $|V(H)| = n$. Then*

$$1. \lambda(H) \geq \frac{|\mathcal{E}(H)|}{n^r} = P_H\left(\frac{1}{n} \cdot \bar{1}\right).$$

Now, let $\bar{x} \geq 0$, $\sum_{i \in [n]} x_i = 1$ be such that $\lambda(H) = P_H(\bar{x})$. Moreover, assume for every such \bar{x} , we have $i \in [n]$, $x_i > 0$. Then,

2. *For all $u, v \in V(H)$ there exists $e \in \mathcal{E}(H)$ containing both u and v , and*

$$3. \frac{\partial}{\partial x_i} P_H(\bar{x}) = r \cdot \lambda(H).$$

Proof. 1. This is immediate.

2. Suppose no edge contains u and v . Then

$$P_H(\bar{x}) = C_u x_u + C_v x_v + b$$

where C_u, C_v, b do not depend on x_u and x_v . As in the graph case, assume $C_u \geq C_v$. Let x' be such that

$$x'_u = x_u + x_v, \quad x'_v = 0, \quad x'_w = x_w \text{ for } w \neq \{u, v\}$$

then $P_H(x') \geq P_H(\bar{x}) = \lambda(H)$, contradicting the conditions on H .

These are typically extremely difficult.

[†] An old and important problem.

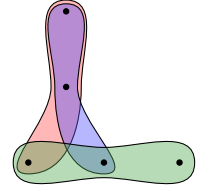


Figure 23: An illustration of T_3 , numbered from top to bottom, left to right. Then D_1 is in red, D_2 in blue, and D_3 in green.

That is, H covers all pairs.

Namely that every extremal \bar{x} has $x_v > 0$.

3. Lagrange multiplier principle: let \bar{x} be a local maximum of $f(\bar{x})$ such that $g(\bar{x}) = 0^\dagger$. Then

$$\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) = \nabla f, \text{ and } \left(\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_n} \right) = \nabla g$$

are colinear. Applied to $f = P_H$ and $g = \sum x_i - 1$, this principle implies

$$C = \frac{\partial P_H}{\partial x_i} = \frac{\partial P_H}{\partial x_j}$$

for all i, j . Then

$$\begin{aligned} \sum_{i=1}^n x_i \frac{\partial P_H}{\partial x_i} &= \sum_{i=1}^n x_i \frac{\partial}{\partial x_i} \sum_{e \in H} \prod_{j \in e} x_j \\ &= \sum_{i=1}^n \sum_{e \in H} x_i \delta_{i \in e} \prod_{j \in e - \{i\}} x_j \\ &= \sum_{e \in H} \sum_{i=1}^n \delta_{i \in e} \prod_{j \in e} x_j \\ &= \sum_{e \in H} r \prod_{j \in e} x_j = r P_H(\bar{x}). \end{aligned}$$

Then

$$C = \sum_{i=1}^n x_i C = r \lambda(H).$$

[†] A constraint.

This argument works for all multilinear polynomials homogeneous of degree r .

□

Define a 3-graph $K_4^- = \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}\}$; see fig. 24.

Lemma 6.2. *Let H be a 3-graph (with at least one edge) with no T_3 subgraph and no K_4^- subgraph, then $\lambda(H) = \frac{1}{27} = \frac{2}{9} \frac{1}{3!}$.*

Proof. We may assume that any vector \bar{x} maximizer of $P_H(\bar{x})$ such that $\bar{x} \geq 0$, $\sum x_i = 1$ has no zero components[‡]. Then by lemma 6.1(2), H covers pairs. In fact, as H contains no T_3 or K_4^- , every pair in H is covered by exactly one edge. This argument is illustrated in fig. 25.

Let $|V(H)| = n$. Then, using lemma 6.1(3), we have

$$3n\lambda(H) = \sum_{i=1}^n \frac{\partial}{\partial x_i} P_H(\bar{x}) = \sum_{\{i,j\} \in [n]^{(2)}} x_i x_j \leq \lambda(K_n) = \frac{1}{2} \frac{n-1}{n}.$$

Then

$$\lambda(H) \leq \frac{1}{6} \frac{n-1}{n^2} \leq \frac{1}{6} \cdot \frac{2}{9} = \frac{1}{27},$$

taking $n = 3$ and using that $\frac{n-1}{n^2}$ is decreasing for $n > 2$. On the other hand, by lemma 6.1(1), any 3-graph H with an edge has

$$\lambda(H) \geq \frac{|\mathcal{E}(H)|}{n^r} \geq \frac{1}{3^3} = \frac{1}{27}.$$

□

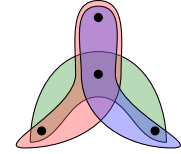


Figure 24: An illustration of K_4^- .

[‡] Otherwise, remove those vertices; this doesn't change $P_H(\bar{x})$ or $\lambda(H)$.

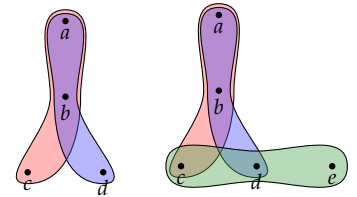


Figure 25: *Left:* A depiction of two edges covering the pair $\{a, b\}$, yielding two distinct points c and d . *Right:* some edge must cover $\{c, d\}$; this edge has a third point e . If $e \in \{a, b\}$, then we have a K_4^- subgraph; otherwise, a T_3 subgraph.

Theorem 6.3 (Erdős and Simonovits 1983). Assume $\mathcal{F} = \{F_1, \dots, F_k\}$ is a family of r -graphs. Then for every ϵ , there exists $\delta > 0$ and n_0 such that if G is an r -graph on $n \geq n_0$ vertices with $|\mathcal{E}(G)| \geq (\pi(\mathcal{F}) + \epsilon) \binom{n}{r}$, then for some $1 \leq i \leq k$, then G contains at least $\delta \binom{n}{|V(F_i)|}$ copies of F_i .

We will apply this result to T_3 before proving it.

Theorem 6.4.

$$\pi(T_3) = \frac{2}{9}.$$

Proof. Suppose $\pi(T_3) = \frac{2}{9} + 2\epsilon$ for some $\epsilon > 0$. Let $\mathcal{F} = \{T_3, K_4^-\}$. Then $\pi(\mathcal{F}) = \frac{2}{9}$ by lemma 6.2. Let δ, n_0 satisfy theorem 6.3 for these \mathcal{F}, ϵ . Let G have $|V(G)| = n > n_0$, $|\mathcal{E}(G)| \geq (\frac{2}{9} + \epsilon) \binom{n}{3}$. We will show that (if n is large enough) G contains T_3 . This will finish the proof.

If not, then by theorem 6.3, G contains $\geq \delta \binom{n}{4}$ copies of K_4^- . Then there exists a pair $a, b \in V(G)$ such that a, b belong to $\geq \frac{\delta \binom{n}{4}}{\binom{n}{2}} \geq \frac{\delta}{12} n^2$ copies of K_4^- in which neither a nor b belong to all three edges[†].

Let one such copy of K_4^- be that of fig. 26. Then $\{a, b\}$ must belong to an edge which contains neither c nor d (otherwise $\{a, b\}$ is only in two edges: $\{a, b, c\}$ and $\{a, b, d\}$). Let $\{a, b, e\}$ be such an edge. Then $\{\{a, c, d\}, \{b, c, d\}, \{a, b, e\}\}$ is T_3 as desired. \square

[†] Why? Pigeonhole argument: we are distributing $\delta \binom{n}{4}$ copies of K_4^- among $\binom{n}{2}$ pairs of vertices; some pair of vertices must end up with at least $\frac{\delta \binom{n}{4}}{\binom{n}{2}}$ copies of K_4^- .

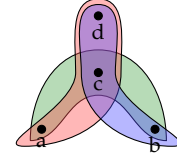


Figure 26: A labelled illustration of K_4^- .

Proof of theorem 6.3. Let us only consider the case $\mathcal{F} = \{F\}$ for simplicity. Let n' be such that every graph G on n' vertices with $|\mathcal{E}(G)| \geq (\pi(\mathcal{F}) + \frac{\epsilon}{2}) \binom{n'}{r}$ contains F ; such an n' exists by the definition of limit. We will choose n_0 appropriately large compared to n' .

Claim. At least $\frac{\epsilon}{2} \binom{n'}{r}$ subgraphs of G on n' vertices have at least $(\pi(F) + \frac{\epsilon}{2}) \binom{n'}{r}$ edges, and thus have a copy of F .

Proof of claim. Suppose not. (Let $V(G) = [n]$). Then

$$|\mathcal{E}(G)| \underbrace{\binom{n-r}{n'-r}}_{\substack{\# \text{ of } n' \text{-element sets} \\ \text{containing an edge}}} = \sum_{X \in [n]^{(n')}} \underbrace{|\mathcal{E}(G|X)|}_{\text{edges in } G \text{ restricted to } X}$$

By our assumption,

$$\begin{aligned} \sum_{X \in [n]^{(n')}} |\mathcal{E}(G|X)| &= \sum_{X: |\mathcal{E}(X)| \geq (\pi(F) + \frac{\epsilon}{2}) \binom{n'}{r}} \mathcal{E}(G|X) + \sum_{X: |\mathcal{E}(X)| < (\pi(F) + \frac{\epsilon}{2}) \binom{n'}{r}} \mathcal{E}(G|X) \\ &\leq \frac{\epsilon}{2} \binom{n}{n'} \underbrace{\binom{n'}{r}}_{\text{maximal \# edges}} + (1 - \frac{\epsilon}{2}) \binom{n}{n'} (\pi(\mathcal{F}) + \frac{\epsilon}{2}) \binom{n'}{r} \end{aligned}$$

On the other hand, $|\mathcal{E}(G)| \geq (\pi(F) + \epsilon) \binom{n}{r}$, so we have

$$(\pi(F) + \epsilon) \binom{n}{r} \binom{n-r}{n'-r} \leq \frac{\epsilon}{2} \binom{n}{n'} \binom{n'}{r} + (1 - \frac{\epsilon}{2}) \binom{n}{n'} (\pi(\mathcal{F}) + \frac{\epsilon}{2}) \binom{n'}{r}.$$

But, by eq. (1), we have $\binom{n}{r}\binom{n-r}{n'-r} = \binom{n}{n'}\binom{n'}{r}$, so

$$\begin{aligned}\pi(F) + \epsilon &\leq \frac{\epsilon}{2} + (1 - \frac{\epsilon}{2})(\pi(F) + \frac{\epsilon}{2}) \\ \pi(F) + \frac{\epsilon}{2} &\leq \pi(F) + \frac{\epsilon}{2} - \frac{\epsilon}{2}\pi(F) - \frac{\epsilon^2}{4} \\ 0 &\leq -\frac{\epsilon}{2}\pi(F) - \frac{\epsilon^2}{4}\end{aligned}$$

which is a contradiction, since $\pi(F) \geq 0$. \blacksquare

So by the choice of n' each of these subgraphs contains F . Suppose G contains k copies of F , and $v = |V(F)|$. Then each of these copies is counted in $\binom{n-v}{n'-v}$ many n' -vertex subgraphs. So, by our claim, $k\binom{n-v}{n'-v} \geq \frac{\epsilon}{2}\binom{n}{n'}$. Then,

$$k \geq \frac{\epsilon}{2} \frac{\binom{n}{n'}}{\binom{n-v}{n'-v}} = \frac{\epsilon}{2} \frac{\binom{n}{v}}{\binom{n'}{v}} = \frac{\epsilon}{2} \binom{n}{v} = \delta \binom{n}{v}$$

where $\delta := \frac{\epsilon}{2} \frac{\binom{n}{v}}{\binom{n'}{v}}$, and we've used eq. (1) again. \square

6.1 Turán density of Fano plane.

The Fano plane F_7 is the 3-graph on vertices $\mathbb{Z}_2^3 - \{(0,0,0)\}$, i.e. triples of 0s and 1s not all zero. We declare $\{x, y, z\}$ to be an edge of F_7 if $x + y + z = 0 \iff x + y = z$; see fig. 27 for an illustration. Note there are seven edges.

Recall that $c : V(H) \rightarrow [k]$ is a k -coloring of a hypergraph H if no edge of H is monochromatic, and $\chi(H)$ is the minimum k such that H admits a k -coloring. We have the following lemma.

Lemma 6.5.

$$\chi(F_7) = 3.$$

Proof. In a random 3-coloring, the probability that an edge is monochromatic is $\frac{1}{9}$, so the expected number of monochromatic edges is $\frac{7}{9} < 1$. So there exists a 3-coloring with no monochromatic edges. So $\chi(H) \leq 3$.

Suppose there exists a 2-coloring of F_7 . Then by pigeonhole, at least four vertices have to receive the same color; say w, x, y, z , say with color 1. Since $\{x, y, x+y\}$ is an edge: $x + y + x + y = 2x + 2y = 0$, we must have $x + y$ be color 2; similarly with $y + z$ and $x + z$. But then $\{x + y, y + z, x + z\}$ is a monochromatic edge, of color 2. \square

If we partition n vertices into two approximately equal halves, we may define a 3-graph by declaring all triples with at least one member in each half to be edges; see fig. 28 for an illustration. There are

Finite projective geometry

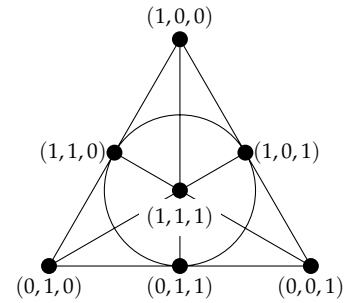


Figure 27: A depiction of the Fano plane F_7 . Each straight line is an edge, and the circle is an edge; each edge contains exactly 3 vertices, since F_7 is a 3-graph.

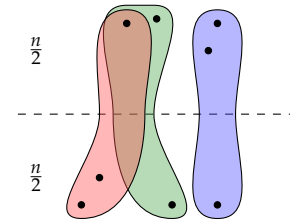


Figure 28: A 3-graph on n vertices formed by partitioning the vertices in two approximately equal parts, and choosing edges as all triples with at least one vertex in each part. For clarity, not all the edges are drawn here.

$\frac{3}{4}\binom{n}{3} + O(n^2)$ edges in this graph. If this graph had a Fano plane subgraph, then we could 2-color the Fano plane, which is impossible by lemma 6.5. Thus, the graph of fig. 28 must not have an F_7 subgraph, so $\pi(F_7) \geq \frac{3}{4}$.

Theorem 6.6 (de Caen and Füredi 2000).

$$\pi(F_7) = \frac{3}{4}.$$

Proof. Suppose not: $\pi(F_7) = \frac{3}{4} + 2\epsilon$ for some $\epsilon > 0$. By lemma 5.7, for every n_0 there exists a 3-graph G on $n \geq n_0$ vertices with no F_7 subgraph such that every vertex of G belongs to $\geq (\frac{3}{4} + \epsilon)\binom{n}{2}$ edges. Moreover, we may assume that G contains $K_4^{(3)}$: four vertices such that every triple forms an edge[†]. We'll take these vertices to be a, b, c, d .

For $x \in V(G)$, let $L_x = \{\{y, z\} : \{x, y, z\} \in \mathcal{E}(G)\}$. This is called the *link graph* of x in G . For x_1, x_2, \dots, x_k we will consider $L_{x_1} \cup L_{x_2} \cup \dots \cup L_{x_k}$ (abusing notation) as a *multigraph*, meaning that we count multiplicity.

Claim. If $\{a, b, c\} \in \mathcal{E}(G)$, then $L_a \cup L_b \cup L_c$ induces at most 15 edges (counting multiplicity) on any four vertices.

Proof of claim. Let x, y, z, w be vertices in $V(G) - \{a, b, c\}$. As G contains no F_7 , there does not exist a partition of the edge set of the complete 2-graph on $\{x, y, z, w\}$ into three subsets M_1, M_2 , and M_3 of size two such that $|M_1 \cap L_a| = 2, |M_2 \cap L_b| = 2$, and $|M_3 \cap L_c| = 2$; see fig. 29.

Suppose $L_a \cup L_b \cup L_c$ has ≥ 16 edges on $\{x, y, z, w\}$, i.e., these link graphs missed at most 2 out of the maximum possible number of edges (which is $3 * \binom{4}{2} = 18$, recalling that we count edges in the link graphs with multiplicity). Let

$$M_1 = \{\{x, y\}, \{z, w\}\}, \quad M_2 = \{\{x, w\}, \{y, z\}\}, \quad M_3 = \{\{x, z\}, \{y, w\}\}$$

as shown in fig. 30. Up to changing the labels, there are two cases.

Case 1: Both missing edges belong to M_1 . That is, either for some $x \in \{a, bc\}$, we have $|L_x \cap M_1| = 0$ or for some $x, y \in \{a, b, c\}$ distinct, we have $|L_x \cap M_1| = 1$ and $|L_y \cap M_1| = 1$.

In any case, M_1 fully intersects with one of the L_x ; say, L_a (that is, $|L_a \cap M_1| = 2$). Then since those were the only two missing edges, we have $|L_b \cap M_2| = |L_c \cap M_3| = 2$.

Case 2: One missing edges belongs to M_1 and one M_2 . That is, $|M_1 \cap L_x| = 1$ and $|M_2 \cap L_y| = 1$, for some $x, y \in \{a, b, c\}$. Since those were the only missing edges, M_1 fully intersects with

[†] If not, every set of four vertices yields at most three edges, which implies G has at most $\frac{3}{4}\binom{n}{4}$ edges in total, contradicting our assumption.

A single link graph can induce $\binom{4}{2} = 6$ edges on four vertices, so our naive bound is 18. Our claim is that this additional structure of being link graphs of G lets us do slightly better.

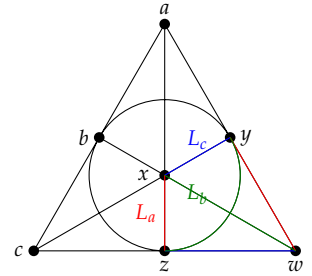


Figure 29: Given an edge $\{a, b, c\}$, we consider any four vertices $x, y, z, w \in L_a \cup L_b \cup L_c$, and on these seven vertices, we construct the Fano plane. Since G contains no F_7 subgraph, the edges drawn here must not all appear in the link graphs.

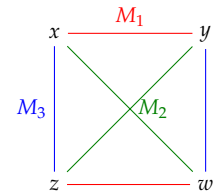


Figure 30: A choice of partition M_1, M_2, M_3 of edges of the complete graph on $\{x, y, z, w\}$, chosen to correspond to fig. 29.

two link graphs, and M_2 with two. So we may choose them to be different, putting, say, $|M_1 \cap L_a| = 2$, $|M_2 \cap L_b| = 2$, and $|M_3 \cap L_c| = 2$. ■

Claim. If a, b, c, d are the vertices of $K_4^{(3)}$, then $L_a \cup L_b \cup L_c \cup L_d$ induces at most 20 edges counted with multiplicities in $V(G) - \{a, b, c, d\}$.

Proof. Apply previous claim to all 4 edges $\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, $\{b, c, d\}$. In total, the resulting multigraph will have at most 60 edges but each L_i (for $i \in \{a, b, c, d\}$) is counted 3 times. ■

Then,

$$|L_a \cup L_b \cup L_c \cup L_d| \geq (3 + \epsilon) \binom{n}{2}$$

If we restrict to $L_a \cup L_b \cup L_c \cup L_d$ to $V(G) - \{a, b, c, d\}$, we still have at least

$$(3 + \epsilon) \binom{n}{2} - 12n \quad (\star f)$$

edges, where we've overcounted edges in $L_a \cup L_b \cup L_c \cup L_d$ using one of $\{a, b, c, d\}$.

We may think of a multigraph with vertex set V as a function $w : V^{(2)} \rightarrow \mathbb{Z}_+$. Let $m(n, k, r)$ be the maximum number of edges (the sum of weights $\sum_{e \in V^{(2)}} w(e)$) in a multigraph on n vertices such that any k vertices induce $\leq r$ edges. We will show that

$$m(n, 4, 20) \leq 3 \binom{n}{2} + n - 2 \quad (\star)$$

for $n \geq 4$. This will finish the proof, as it contradicts $(\star f)$ for large n . We will assume

$$m(n, 3, 10) \leq 3 \binom{n}{2} + n - 2. \quad (\star \star)$$

The proof of this is left as an exercise. Let us prove (\star) by induction. Base case:

$$m(4, 4, 20) \leq 3 \cdot 6 + 4 - 2 = 20.$$

Induction step: Using $(\star \star)$, we assume $\{a, b, c\} \subset V$ induces $k \geq 11$ edges. Let $w(x) = \sum_{e \ni x} w(e)$ for $x \in \{a, b, c\}$. Then

$$w(a) + w(b) + w(c) = \underbrace{2k}_{\text{weight of edges with two of } \{a, b, c\}} + \sum_{z \in V - \{a, b, c\}} (w(za) + w(zb) + w(zc))$$

Since every four vertices span at most 20 edges, for each $z \in V - \{a, b, c\}$ we have $w(za) + w(zb) + w(zc) \leq 20 - k$. Then

$$\begin{aligned} w(a) + w(b) + w(c) &\leq 2k + (n - 3)(20 - k) \\ &= 20(n - 3) - k(n - 5) \\ &\leq 20(n - 3) - 11(n - 5) = 9n - 5 \end{aligned}$$

Note in the following, we are considering 2-graphs.

WLOG, $w(a) \leq 3n - 2$ by pigeonhole. By the induction hypothesis, $V - \{a\}$ induces total weight at most

$$3\binom{n-1}{2} + (n-1) - 2.$$

So the total number of edges is at most

$$3\binom{n-1}{2} + (n-1) - 2 + (3n-2) \leq 3\binom{n}{2} + n - 2. \quad \square$$

References for Section 6.

- de Caen, D. and Z. Füredi (2000). “The Maximum Size of 3-Uniform Hypergraphs Not Containing a Fano Plane”. In: *Journal of Combinatorial Theory, Series B* 78.2, pp. 274–276. ISSN: 0095-8956. DOI: <http://dx.doi.org/10.1006/jctb.1999.1938> (cit. on p. 46).
- Erdős, P. and M. Simonovits (1983). “Supersaturated graphs and hypergraphs”. In: *Combinatorica* 3.2, pp. 181–192. ISSN: 1439-6912. DOI: [10.1007/BF02579292](https://doi.org/10.1007/BF02579292) (cit. on p. 44).

Remark. In particular, $R(k, k) \leq \binom{2k-2}{k-1} \sim \frac{4^k}{\sqrt{k}}$. We in fact have $R(3, 3) = 6$, $R(4, 4) = 18$, and $43 \leq R(5, 5) \leq 49$.

Theorem 7.3 (Erdős 1947). $R(k, k) \geq 2^{k/2}$ for $k \geq 2$.

Proof. Let $n = \lfloor 2^{k/2} \rfloor$ and let $c : [n]^{(2)} \rightarrow \{R, B\}$ be chosen uniformly randomly. That is, the color of every edge in $[n]^{(2)}$ is chosen to be R with probability $1/2$ or B with probability $1/2$, independently of the other edges. Let Z be equal to the number of sets $X \subset [n]$, $|X| = k$, such that c restricted to $X^{(2)}$ is monochromatic. It is enough to show $\mathbb{E}[Z] < 1$. For fixed X as above,

$$\Pr[c \text{ restricted to } X \text{ is monochromatic}] = \frac{1}{2^{\binom{k}{2}-1}}.$$

Then

$$\begin{aligned} \mathbb{E}[Z] &= \binom{n}{k} 2^{1-\binom{k}{2}} \leq \frac{n^k}{k!} 2^{1-\frac{k(k-1)}{2}} \\ &\leq \frac{2^{k^2/2} \cdot 2^{1+(k/2)-k^2/2}}{k!} = \frac{2^{1+k/2}}{k!} < 1 \end{aligned}$$

for $k \geq 3$. One may show $R(2, 2) = 2$, yielding the case $k = 2$. \square

The state of the art bounds are

$$(1 + O(1)) \frac{\sqrt{2k}}{e} 2^{k/2} \leq R(k, k) \leq k^{-\frac{c \log k}{\log \log k}} 4^k.$$

NEXT, IF n IS SUFFICIENTLY LARGE with respect to k and $c : [n] \rightarrow \{R, B\}$, then $[n]$ contains a monochromatic *arithmetic progression* of length k , that is a set

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

for some $d > 0$. We may write this as $a + [0, k - 1]d$, where $[0, k - 1] = \{j \in \mathbb{Z} : 0 \leq j \leq k - 1\}$ and the addition and multiplication is elementwise. The *Van der Waerden number* $W(k, r)$ is the minimum n such that if $[n]$ is colored in r colors one can always find a monochromatic arithmetic progression of length k .

Theorem 7.4 (van der Waerden 1927). $W(k, r)$ exists for all r, k .

The proof will follow from lemma 7.5.

Example. • $W(2, r) = r + 1$

• $W(k, 1) = k$,

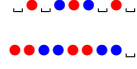
Since then there must be exist some coloring with $Z < 1$, i.e., $Z = 0$.

Lower: Spencer 1975, upper: Conlon 2009.

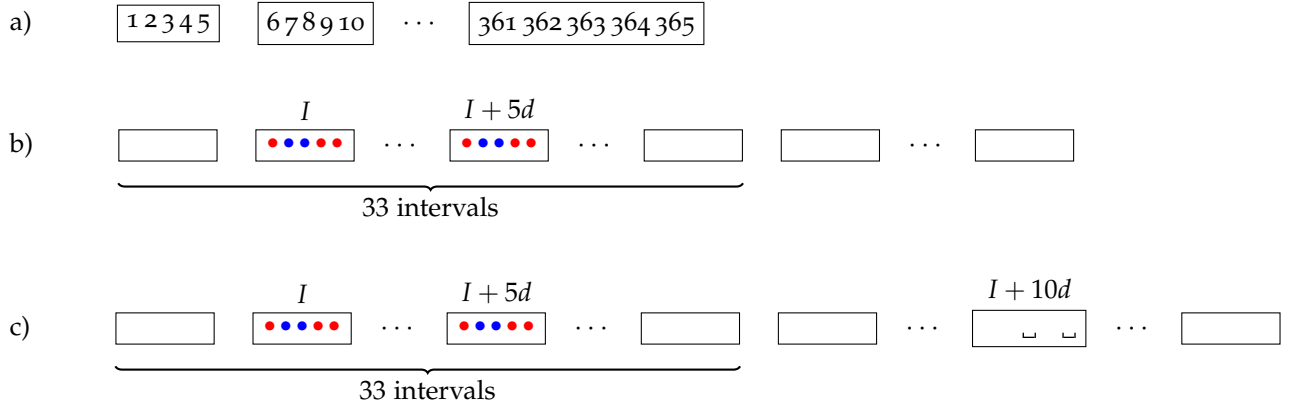
Note that here we are discussing colorings of $[n]$, not $[n]^{(2)}$; that is, vertex colorings, not edge colorings.

by pigeonhole

- What about $W(3,2)$? Well, $\bullet \bullet \bullet \bullet \bullet \bullet$ shows $W(3,2) > 8$. On the other hand, assuming the 5th slot is red, wlog, we reduce to two cases:



Let's outline the general argument; this will give a bound $W(3,2) \leq 5 \times 65 = 325$. Color $[325]$ and divide it into 65 intervals of length 5.



Consider the first $2^5 + 1 = 33$ of them. Two of them will be colored exactly the same by pigeonhole; say interval I and interval $I + 5d$. If $I = [a, a + 5]$, then $\{a, a + d', a + 2d'\}$ is an arithmetic progression for $d' \in \{1, 2\}$. Now, if either of these progressions are monochromatic, then we are done. Otherwise, say we have $\{a, a + d', a + 2d'\}^\dagger$, for $d' \in \{1, 2\}$. Now, we consider the interval $I + 10d$. If $a + 2d' + 10d$ is red, then $\{a + 2d', a + 2d' + 5d, a + 2d' + 10d\}$ is monochromatic red. Otherwise $a, a + d' + 5d, a + 2d' + 10d$ is monochromatic blue.

A *polychromatic m -tuple* of arithmetic progression of length k is a set $A = \{a + [0, k]d_1\} \cup \{a + [0, k]d_2\} \cup \dots \cup \{a + [0, k]d_m\}$ such that $a + [1, k]d_i$ is monochromatic for all i , and all of these m progressions are of different colors (see fig. 32 for an example).

Lemma 7.5. *If $W(k-1; r)$ exists for every r then for every m there exists $W(k, m, r) = n$ such that in any r -coloring of $[n]$ there exists a monochromatic a.p. of length k or a polychromatic m -tuple of a.p. of length $k-1$.*

Remark. With this result, theorem 7.4 follows by induction on k , using that $W(k; r) \leq W(k, r+1, r)$, since there cannot exist a polychromatic $r+1$ -tuple when there are only r colors.

Figure 31: a) We subdivide $[325]$ into 65 intervals of length 5. b) By the pigeonhole principle, two of the first $2^5 + 1 = 33$ intervals have same coloring. If I is the first interval; then for some $d > 0$, the second is $I + 5d$. c) To complete the proof, we consider $I + 10d$, as described in the text.

[†] If a and $a + d'$ are different colors, then choose the other d' .

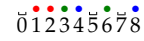


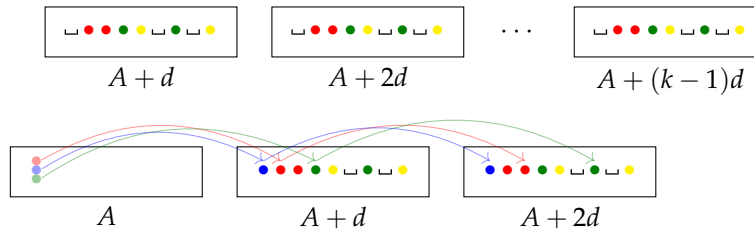
Figure 32: With the coloring shown, $\{0 + [0, 2]\} \cup \{0 + [0, 2] \cdot 3\} \cup \{0 + [0, 2] \cdot 4\}$ is a polychromatic 3-tuple of length 2.

Proof by induction on m . Base case: $m = 1$. Then in a set of size $W(k-1, r)$ we may find a monochrome $(k-1)$ -a.p., say $a + d[1, k]$; then for a polychromatic 1-tuple of a.p. of length k , we simply need to add the point a . We may ensure this by taking a second copy of $W(k-1, r)$ to the left, yielding the bound $W(k, 1, r) \leq 2W(k-1, r)$ (see fig. 33 for an illustration).

To show the induction step, we will first prove the following result.

Claim. If $A + d, A + 2d, \dots, A + (k-1)d$ are *identically colored*[†] polychromatic m -tuples of a.p. of length $(k-1)$ then $A \cup (A + d) \cup \dots \cup (A + (k-1)d)$ contains a polychromatic $(m+1)$ -tuple of a.p. of length $k-1$, or a monochromatic a.p. of length k .

Proof.



Write $A + d = \bigcup_{i=1}^m \{a + d + d_i[0, k-1]\}$ where each $a + d + d_i[0, k-1]$ is monochromatic and differently colored from $a + d + d_j[0, k-1]$ for $i \neq j$. In this language, the identical coloring assumption is that $\{a + \ell d + s d_i : \ell \in [k-1]\}$ is monochromatic for each $s \in [0, k-1]$. In particular,

$$P' := \{a + d, a + 2d, \dots, a + (k-1)d\}$$

is a monochromatic $(k-1)$ -a.p. If it is the same as color as $a + d + d_i$ for some i , then

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

is a monochromatic k -a.p. Otherwise, consider

$$P'_i := \{a + d + d_i, a + 2d + 2d_i, \dots, a + (k-1)d + (k-1)d_i\}$$

which is an a.p. of the same color as

$$\{a + d + d_i, a + d + 2d_i, \dots, a + d + (k-1)d_i\}.$$

and thus a different color from P' . Then $P' \cup P'_1 \cup \dots \cup P'_m \cup \{a\}$ is a polychromatic $(m+1)$ -tuple. ■

This is the meat of the proof.

Now let $M = W(k, m, r)$ and $N = W(k-1, r^M)$. We will show that $W(k, m+1, r) \leq 2MN$.

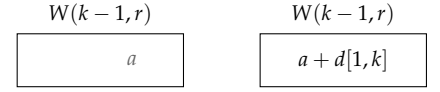


Figure 33: Depiction showing $W(k, 1, r) \leq 2W(k-1, r)$ by finding a monochrome $(k-1)$ -a.p. $a + d[1, k]$ in a set of size $W(k-1, r)$ and adding in the point a , which may be to the left of the original set of size $W(k-1, r)$.

[†] For each $x \in A + d$, we have that $x, x + d, \dots, x + (k-2)d$ all have the same color.

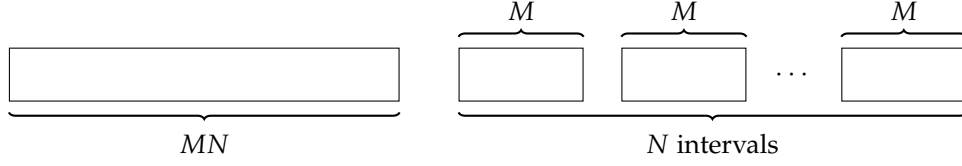


Figure 35: We think of coloring each interval of size M on the right side into r colors as assigning the entire interval one of r^M colors.

We divide $2MN$ into an interval of size MN followed by N intervals of size M , as shown in fig. 35. By choice of N , there are intervals $(k-1)$ intervals on the right side of the form $I, I+d, I+2d, \dots, I+(k-2)d$ which are identically colored.

By the choice of M , we may assume I contains a polychromatic m -tuple of a.p. of length $(k-1)$ which we will call $A+d$. The first interval of size MN serves to include A . The induction step is finished by the claim. \square

The proof yields very poor bounds for $W(k; r)$. The current best bounds are

$$(1 + o(1)) \frac{r^k}{\text{erk}} < W(k; r) \leq 2^{2^{2^{k+9}}}.$$

LHS: folklore with a randomized construction. RHS: Gowers 2001.

It's conjectured that the upper bound can be improved substantially:

Conjecture. $W(k; 2) \leq 2^{k^2}$.

Let us proceed to the Hales-Jewett theorem. It may be informally stated as the following: in a $\overbrace{t \times t \cdots \times t}^{d \text{ dimensional}}$ game of tic-tac-toe with r players, a draw is impossible as long as d is large enough compared to r and t . Let A be a finite alphabet of size t , typically $[0, t-1]$. Then A^d is the set of ordered d -tuples of elements of A , or words of length d in the alphabet A .

In tic-tac-toe, $A = \{0, 1, 2\}$, and $d = 2$ (see fig. 36). We think of each player having a color; a coloring of $\{0, 1, 2\}^2$ by two colors then corresponds to the moves made by both the players. A draw is impossible if in any coloring of $\{0, 1, 2\}^2$, there is a monochromatic “3-in-a-row”, i.e., row, column, or diagonal.

Moving back to the general development, a *root* τ is a word of length d in the alphabet $A \cup \{\star\}$, where \star is a symbol not in A , which contains at least one \star . For a root τ and $a \in A$, the word $\tau(a)$ is obtained by substituting a instead of \star everywhere in τ . A *combinatorial line* in A^d is a set $L_\tau := \{\tau(a) : a \in A\}$ where τ is a root of length d . See fig. 37 for examples and nonexamples of combinatorial lines in tic-tac-toe. With these definitions, we may formulate the Hales-Jewett theorem as follows.

2	02	12	22
1	01	11	21
0	00	10	20
	0	1	2

Figure 36: A depiction of words in tic-tac-toe: elements of $\{0, 1, 2\}^2$.

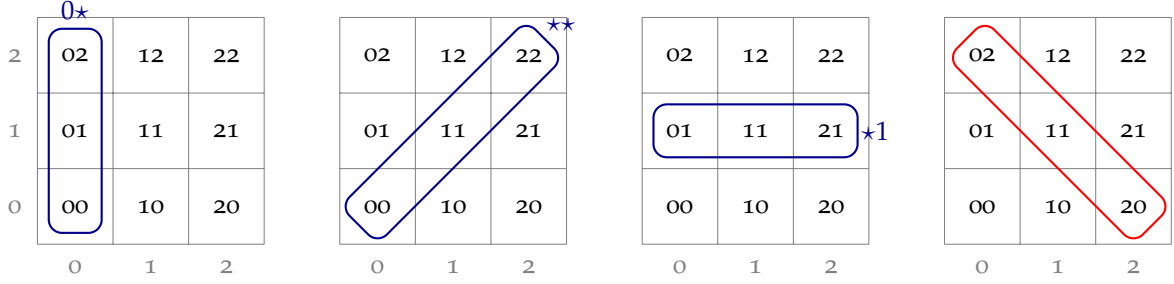


Figure 37: Three examples of combinatorial lines in $\{0, 1, 2\}^2$, followed by a nonexample. From left to right, combinatorial lines corresponding to roots $\tau_1 = 0*$, $\tau_2 = **$, and $\tau_3 = *1$, respectively, followed by the set $\{02, 11, 20\}$ in red which is not a combinatorial line. So in tic-tac-toe, not every “3-in-a-row” is a combinatorial line, but every combinatorial line is a “3-in-a-row”, which is enough to show that draws are impossible if we can always find a monochromatic combinatorial line.

Theorem 7.6 (Hales and Jewett 1963). *For every r and t , there exists $d = \text{HJ}(t, r)$ such that if A is an alphabet with $|A| = t$ and A^d is colored in r colors, then there exists a monochromatic combinatorial line.*

Remark. The same is true for every $d' \geq d$.

Before proving the Hales-Jewett theorem, we’ll discuss an application. Let $V \subset \mathbb{Z}^d$ be a finite collection of vectors. U is called a homothetic copy of $V = \{v_1, v_2, \dots, v_t\}$ if $U = u + \lambda V = \{u + \lambda v_1, u + \lambda v_2, \dots, u + \lambda v_t\}$ for some $u \in \mathbb{Z}^d$ and $\lambda \in \mathbb{Z}$. Homothetic copies are a generalization of arithmetic progressions: for $d = 1$ and $V = \{0, 1, \dots, k-1\}$, a homothetic copy of V is exactly an arithmetic progression of length k . See fig. 38 for a two dimensional example.

Theorem 7.7 (Gallai (late 1930s), Wit 1952). *Let \mathbb{Z}^d be colored in r colors and let $V \subset \mathbb{Z}^d$ be finite. Then there exists a monochromatic homothetic copy of V .*

Remark. In fact, one can replace \mathbb{Z}^d by $[n]^d$ where n depends on V and r , yielding an analog of van der Waerden’s theorem.

Proof. The proof will be based on the HJ theorem.

We will take an appropriately large hypercube and project it into \mathbb{Z}^d such that a combinatorial line in the hypercube projects to a homothetic copy of V . Let $V = \{v_1, \dots, v_t\} =: A$. Let n be such that every r -coloring of A^n contains a monochromatic combinatorial line.

Define $f : A^n \rightarrow \mathbb{Z}^d$ by $f((a_1, a_2, \dots, a_n)) = \sum_{i=1}^n a_i$ where $a_i \in V$. If χ is the coloring of \mathbb{Z}^d into r -colors, then we can define $\chi' : A^n \rightarrow [r]$ by $\chi'(a) = \chi(f(a))$. By the choice of n , there exists a root τ such that $\tau(v_1), \tau(v_2), \dots, \tau(v_t)$ all receive the same color. Then $\{f(\tau(v_1)), f(\tau(v_2)), \dots, f(\tau(v_t))\}$ is a homothetic copy of V , as follows. If $\tau = a_1 a_2 \dots a_n$ then set $u := \sum_{i: a_i \neq * } a_i$, and λ the number of $*$ ’s in τ . Then $f(\tau(v_i)) = u + \lambda v_i$.

For example, consider $V = \{v_1, v_2, v_3\}$ and $\tau = v_1 v_2 * v_1 *$. Then we’re claiming $\{f(\tau(v_1)), f(\tau(v_2)), f(\tau(v_3))\}$ is a homothetic copy of

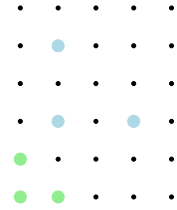


Figure 38: An example of homothetic copies in \mathbb{Z}^2 : the green dots are $V = \{(0,0), (0,1), (1,0)\}$, and the blue dots are $U = u + \lambda V$ for $u = (1,2)$ and $\lambda = 2$.

According to Soifer (2010, page 22), this theorem was communicated from Gallai to Rado in the late 1930s, and was published in 1943 (Rado 1943). The theorem was independently proven by Wit in 1952.

V. E.g.,

$$f(\tau(v_1)) = v_1 + v_2 + v_1 + v_1 + v_1 = 4v_1 + v_2 = \underbrace{v_1 + v_2 + v_1}_u + \underbrace{2}_{\lambda} v_1. \quad \square$$

Proof of theorem 7.6. Let $n = \text{HJ}(t, r)$. We will prove its existence by induction on t for fixed r . First, $\text{HJ}(1, r) = 1$, since combinatorial lines of length 1 are certainly monochromatic.

Now, the induction step. Assume $n = \text{HJ}(t-1, r)$. Let $n \ll N_1 \ll N_2 \ll \dots \ll N_n$. Specifically, $N_1 = r^{t^n}$ and $N_i = r^{t^n + \sum_{j=1}^{i-1} N_j}$ for $i \geq 2$, and $N = \sum_{i=1}^n N_i$. We will show that $\text{HJ}(r, t) \leq N$, i.e. if $\chi : A^N \rightarrow [r]$, then χ contains a monochromatic combinatorial line.

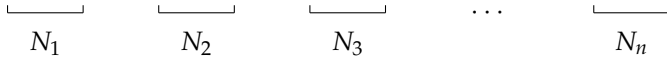


Figure 39: A depiction of N_1, \dots, N_n . We aim to show $\text{HJ}(t, r) = N := \sum_{i=1}^n N_i$.

We will say $a, b \in A^n$ are *neighbors* if they differ in only one position and one of them has symbol 0 in this position and the other has the symbol 1; that is, if $a = a_1 a_2 \dots a_{i-1} 0 a_{i+1} \dots a_n$, and $b = a_1 a_2 \dots a_{i-1} 1 a_{i+1} \dots a_n$ for some $1 \leq i \leq n$.

Let τ be a root of length N such that $\tau = \tau_1 \dots \tau_n$ and τ_i is a root of length N_i . For $a \in A^n$, $a = a_1 \dots a_n$, define

$$\tau(a) = \tau_1(a_1) \tau_2(a_2) \dots \tau_n(a_n).$$

For example, suppose we have a root $\tau = \overbrace{\star}^{\tau_1} \overbrace{2\star}^{\tau_2} \overbrace{\star 3\star}^{\tau_3}$. Then $\tau(012) = 021232$. Given $\tau = \tau_1 \tau_2 \dots \tau_n$ as above, define $\chi_\tau : A^n \rightarrow [r]$ by $\chi_\tau(a) = \chi(\tau(a))$. Now, we're not guaranteed that we have monochromatic combinatorial lines in A^n , but we will try to compress to A^{n-1} where we do have such lines.

Claim. There exist roots $\tau_1, \tau_2, \dots, \tau_n$ such that τ_i has length N_i , and, if $\tau = \tau_1 \dots \tau_n$, then

$$\chi_\tau(a) = \chi_\tau(b)$$

for any pair of neighbors a and b in A^n .

We accept this claim for now to finish the proof. Define χ'_τ as the restriction $\chi'_\tau : (A - \{0\})^n \rightarrow [r]$ of χ_τ . Then χ'_τ contains a monochromatic combinatorial line. That is, there exists $v = v_1 v_2 \dots v_n$ with $v_i \in (A - \{0\}) \cup \{\star\}$ such that $v(1), v(2), \dots, v(t-1)$ all have the same color. We wish to show that the line corresponding to $\tau(v)$ is monochromatic in $\chi : A^N \rightarrow [r]$, where we write $\tau(v) := \tau_1(v_1) \dots \tau_n(v_n)$ with $\tau_i(\star) = \tau_i$. That is, we want $\tau(v(0))$,

If $v = (12\star)$, then $\tau(v) = 122\star 3\star$, using our example τ from earlier.

$\tau(\nu(1)), \dots, \tau(\nu(t-1))$ to receive the same color in χ . So we want $\chi_\tau(\nu(a))$ to be the same for $a = 0, 1, \dots, t-1$.

But $\chi_\tau(\nu(1)) = \dots = \chi_\tau(\nu(t-1))$ by the choice of ν . So it remains to show that $\chi_\tau(\nu(0)) = \chi_\tau(\nu(1))$. While $\nu(0)$ and $\nu(1)$ may differ in several positions, in each position one has 1 and the other has 0, so we may change them one at a time using that χ_τ acts the same on neighbors, until we see they have the same coloring.

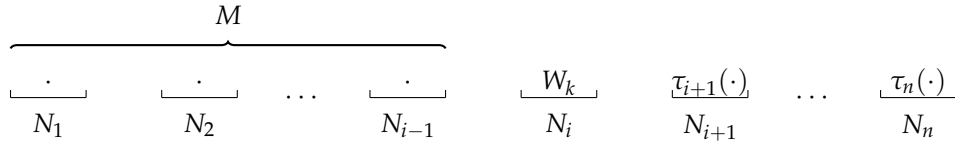
Thus, it remains to prove the claim. We will construct the roots τ_1, \dots, τ_n in reverse order. Suppose $\tau_{i+1}, \dots, \tau_n$ are constructed. For $0 \leq k \leq N_i$, let

$$W_k = \underbrace{0 \dots 0}_k \underbrace{1 \dots 1}_{N_i - k} \in A^{N_i}$$

Let $M = \sum_{j=1}^{i-1} N_j$. Define a coloring $\chi_k : A^{M+n-i} \rightarrow [r]$ as

$$\chi_k(x_1 \dots x_M y_{i+1} \dots y_n) = \chi(x_1 \dots x_M W_k \tau_{i+1}(y_{i+1}) \dots \tau_n(y_n)).$$

See fig. 40 for a depiction of this definition. Now, $N_i > t^{M+n-i}$, so



there exist k, ℓ such that $\chi_k = \chi_\ell$, by the pigeonhole principle. WLOG $k < \ell$. Let $\tau_i = \underbrace{0 \dots 0}_k \underbrace{\star \dots \star}_{\ell-k} \underbrace{1 \dots 1}_{N_i - \ell}$. Then $\tau_i(0) = W_\ell$ and $\tau_i(1) = W_k$.

Now, let's show that the resulting $\tau = \tau_1 \dots \tau_n$ satisfies the claim. Suppose $a = a_1 a_2 \dots a_{i-1} 0 a_{i+1} \dots a_n$, and $b = a_1 a_2 \dots a_{i-1} 1 a_{i+1} \dots a_n$. Then

$$\begin{aligned} \chi_\tau(a) &= \chi(\tau_1(a_1) \tau_2(a_2) \dots \tau_{i-1}(a_{i-1}) W_\ell \tau_{i+1}(a_{i+1}) \dots \tau_n(a_n)) \\ &= \chi_\ell(\tau_1(a_1) \tau_2(a_2) \dots \tau_{i-1}(a_{i-1}) W_\ell a_{i+1} \dots a_n) \\ &= \chi_k(\tau_1(a_1) \tau_2(a_2) \dots \tau_{i-1}(a_{i-1}) a_{i+1} \dots a_n) \\ &= \chi_\tau(b). \end{aligned}$$

□

Theorem 7.8 (Density Hales-Jewett theorem, Furstenberg and Katznelson 1991). *For every $\epsilon > 0$ and each t , there exist n such that if $|A| = t$ and $Z \subset A^n$ with $|Z| \geq \epsilon t^n$ then Z contains a combinatorial line.*

Remark. This theorem implies theorem 7.6. If we consider an r -coloring as a partition Z_1, \dots, Z_r of A^n , then we have some i such that $|Z_i| \geq \frac{1}{r} t^n$ by the pigeonhole principle. Then theorem 7.8 with $\epsilon = \frac{1}{r}$ implies that Z_i contains combinatorial line, proving theorem 7.6.

The case τ_n is similar to the generic step we consider here.

Figure 40: An illustration of how χ_k acts on words in A^{M+n-i} . First, we write such a word as $x_1 \dots x_M y_{i+1} \dots y_n$ where $x_j \in A^{N_j}$ (for $j \in [i-1]$) and $y_j \in A$ (for $j = i+1, \dots, n$). Then, χ_k acts on such a word by creating a word of size A^N as follows, which it plugs into χ . First, x_1, \dots, x_M are not modified (visualized by empty slots of the appropriate size in the figure). Then the word $W_k \in A^{N_i}$ is concatenated, followed by $\tau_j(y_j) \in A^{N_j}$ for $j = i+1, \dots, n$.

Theorem 7.9 (Szemerédi 1975). $\forall \epsilon > 0$ and $\forall k, \exists N > 0$ such that if $A \subset [N]$ with $|A| \geq \epsilon N$, then A contains an arithmetic progression of length k .

Theorem 7.10 (Green and Tao 2008). For every k , the set of primes contain arithmetic progressions of length k .

Since the density of primes goes to zero, the density theorem above does not help.

If we wanted to formulate a density version of Ramsey's theorem, how it would it go? For every t and $\epsilon > 0$, there exists N such that if G is a graph on N vertices and $|G| \geq \epsilon \binom{N}{2}$, then G contains K_t . But this is equivalent to $\pi(K_t) = 0$, which is false for $t \geq 3$.

LET US move on. Consider a finite sequence $A = (a_1, \dots, a_n)$. Then we say $(a_{i_1}, a_{i_2}, \dots, a_{i_k})$ is an *increasing subsequence* of A if $i_1 < i_2 < \dots < i_k$ and $a_{i_1} \leq a_{i_2} \leq \dots \leq a_{i_k}$. Likewise, $(a_{i_1}, a_{i_2}, \dots, a_{i_k})$ is a *decreasing subsequence* of A if $i_1 < i_2 < \dots < i_k$ and $a_{i_1} \geq a_{i_2} \geq \dots \geq a_{i_k}$.

Problem. For all k , does there exist a number $f(k)$ such that every sequence of length at least $f(k)$ contains an increasing or decreasing subsequence of length k ?

Indeed, such an $f(k)$ exists: $f(k) \leq R(k, k)$. Given a sequence $(a_n)_{n=1}^{R(k,k)}$, we can create a coloring on $[R(k, k)]$ as follows: given an edge $\{i, j\} \in [R(k, k)]^{(2)}$ with $i < j$, we color $\{i, j\}$ red if $a_i \leq a_j$, and blue otherwise. Then by the definition of $R(k, k)$, we are guaranteed a monochromatic complete graph on k vertices. This yields an increasing subsequence if the monochromatic K_k is red, and decreasing if the K_k is blue. This yields an exponential bound on $f(k)$. But we can do better.

Theorem 7.11 (Erdős and Szekeres 1935). For all $k, \ell \geq 2$ any sequence of length at least $(k-1)(\ell-1) + 1$ contains an increasing subsequence of length k or a decreasing subsequence of length ℓ .

Remark. In the case $k = \ell$, this is a quadratic bound on $f(k)$.

Proof by induction on $k + \ell$. If $k = 2$, say, then any sequence of length ℓ is decreasing or not. For the induction step, let $n = (k-1)(\ell-1) + 1$, $A = (a_1, \dots, a_n)$ be a sequence, and Z be the set of last elements of increasing subsequences of length $k-1$ in A . Then, $Z = (a_{i_1}, a_{i_2}, \dots, a_{i_z})$ with $i_1 < \dots < i_z$, where $z = |Z|$. For contradiction, assume A violates the claim.

Then $A - Z$ contains no increasing subsequences of length $k-1$ and no decreasing subsequences of length ℓ . So $|A - Z| \leq (k-2)(\ell-1)$. Then $|Z| \geq (k-1)(\ell-1) + 1 - (k-2)(\ell-1) \geq \ell$. So if Z is decreasing we are done. If not, then there exists s such that

$a_{i_s} \leq a_{i_{s+1}}$. Then let $(b_{i_1}, b_{i_2}, \dots, b_{i_{k-1}} = a_{i_s})$ be an increasing sequence ending in a_{i_s} . Then this sequence extends to a length k sequence by adding $a_{i_{s+1}}$. \square

Let us consider a pigeonhole proof of this result.

Proof. As before, let $A = (a_1, \dots, a_n)$ be a sequence. For every $s \in [n]$, we will associate to a_s two numbers:

- i_s , the length of the longest increasing subsequence ending in a_s ,
- d_s , the length of the longest decreasing subsequence ending in a_s .

If A contains no subsequence we need then $1 \leq i_s \leq k-1$ and $q \leq d_s \leq \ell-1$ so there are $(k-1)(\ell-1)$ possible pairs.

So there exist $1 \leq s < t \leq n$ such that $(i_s, d_s) = (i_t, d_t)$ by the pigeonhole principle. Suppose, by symmetry, that $a_s \leq a_t$. Then an increasing subsequence of length i_s ending in a_s can be extended to a subsequence of length i_{s+1} ending in a_t . This is a contradiction to $i_t = i_s$. \square

References for Section 7.

- Conlon, D. (2009). “A new upper bound for diagonal Ramsey numbers”. In: *Annals of Mathematics*, pp. 941–960 (cit. on p. 50).
- Furstenberg, H. and Y. Katznelson (1991). “A density version of the Hales-Jewett theorem”. In: *Journal d’Analyse Mathématique* 57.1, pp. 64–119 (cit. on p. 56).
- Gowers, W. T. (2001). “A new proof of Szemerédi’s theorem”. In: *Geometric & Functional Analysis GAFA* 11.3, pp. 465–588. ISSN: 1420-8970. DOI: [10.1007/s00039-001-0332-9](https://doi.org/10.1007/s00039-001-0332-9) (cit. on p. 53).
- Green, B. and T. Tao (2008). “The primes contain arbitrarily long arithmetic progressions”. In: *Annals of Mathematics* 167-2, pp. 481–547 (cit. on p. 57).
- Hales, A. W. and R. I. Jewett (1963). “Regularity and positional games”. In: *Transactions of the American Mathematical Society* 106.2, pp. 222–229 (cit. on p. 54).
- Erdős, P. (1947). “Some remarks on the theory of graphs”. In: *Bulletin of the American Mathematical Society* 53.4, pp. 292–294 (cit. on p. 50).
- Erdős, P. and G. Szekeres (1935). “A combinatorial problem in geometry”. In: *Compositio Mathematica* 2, pp. 463–470 (cit. on pp. 49, 57, 70).
- Rado, R. (1943). “Note on combinatorial analysis”. In: *Proceedings of the London Mathematical Society* 48.2, pp. 122–160 (cit. on p. 54).
- Ramsey, F.P. (1930). “On a Problem of Formal Logic”. In: *Proceedings of the London Mathematical Society* 2.1, pp. 264–286 (cit. on p. 49).
- Soifer, A., ed. (2010). *Ramsey Theory: Yesterday, Today, and Tomorrow*. Vol. 285. Progress in Mathematics. Birkhäuser Boston. ISBN: 0817680918. DOI: [10.1007/978-0-8176-8092-3](https://doi.org/10.1007/978-0-8176-8092-3) (cit. on p. 54).
- Spencer, J. (1975). “Ramsey’s theorem—A new lower bound”. In: *Journal of Combinatorial Theory, Series A* 18.1, pp. 108–115. ISSN: 0097-3165. DOI: [http://dx.doi.org/10.1016/0097-3165\(75\)90071-0](http://dx.doi.org/10.1016/0097-3165(75)90071-0) (cit. on p. 50).

- Szemerédi, E. (1975). "On sets of integers containing k elements in arithmetic progression". In: *Acta Arithmetica* 27.1, pp. 199–245 (cit. on p. 57).
- van der Waerden, B. L. (1927). "Beweis einer Baudetschen Vermutung." German. In: *Nieuw Arch. Wiskd., II. Ser.* 15, pp. 212–216. ISSN: 0028-9825 (cit. on p. 50).
- Wit, Ernst (1952). "Ein kombinatorischer Satz der Elementargeometrie". In: *Mathematische Nachrichten* 6.5, pp. 261–262. DOI: [10.1002/mana.19520060502](https://doi.org/10.1002/mana.19520060502) (cit. on p. 54).

8 Convexity

We will consider finite collections of points in Euclidean space, \mathbb{R}^d . Let's recall some definitions.

Linear space: $L \subset \mathbb{R}^d$ is a *linear space* if it is closed under addition and multiplication by scalars.

Linear dependence: We say v_1, v_2, \dots, v_n are *linearly dependent* if there exist $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$ not all zero such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0.$$

Linear hull: The *linear hull* $\langle S \rangle$ of $S \subset \mathbb{R}^d$ is the smallest linear subspace containing S , i.e. $\langle S \rangle = \bigcap_{L \supset S} L$ where the intersection is taken over linear subspaces L . Equivalently,

$$\langle S \rangle = \{ \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n : \forall n, v_1, \dots, v_n \in S, \alpha_1, \dots, \alpha_n \in \mathbb{R} \}.$$

Note that we may take $n = d$.

Affine subspace: An *affine subspace* is a subset of \mathbb{R}^d of the form $v + L$ where L is a linear subspace and $v \in \mathbb{R}^d$.

Affine hull: The *affine hull* $\text{aff}(S)$ of $S \subset \mathbb{R}^d$ is the intersection of all affine subspaces containing S . Equivalently,

$$\text{aff}(S) = \{ \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n : \sum_i \alpha_i = 1, v_1, \dots, v_n \in S \}.$$

Note: if $w \in v + L$, then $w - v \in L$, so $w + L = v + L$.

Affine dependence: A set of vectors $\{v_1, \dots, v_n\}$ are *affinely dependent* if

$$\alpha_1 v_1 + \dots + \alpha_n v_n = 0$$

for some $\sum_{i=1}^n \alpha_i = 0$ with some $\alpha_i \neq 0$. This is equivalent to one of the vectors belongs to the affine hull of the other vectors.

Remark. The maximum number of affinely independent vectors in \mathbb{R}^d is $d + 1$. Why? Any $d + 2$ vectors in \mathbb{R}^d are affinely dependent. Suppose we have v_1, \dots, v_{d+2} . Then WLOG v_1, \dots, v_d form a basis of \mathbb{R}^d . Then $\alpha_1 v_1 + \dots + \alpha_d v_d = v_{d+1}$ for some choice of α_i 's.

Likewise, $\beta_1 v_1 + \dots + \beta_d v_d = v_{d+2}$. Let $\alpha = \sum_i \alpha_i$ and $\beta = \sum_i \beta_i$. So

$$\alpha_1 v_1 + \dots + \alpha_d v_d - v_{d+1} = 0$$

$$\beta_1 v_1 + \dots + \beta_d v_d - v_{d+2} = 0.$$

If $\alpha = 1$ or $\beta = 1$, then we are done. Otherwise,

$$(\beta - 1)(\alpha_1 v_1 + \dots + \alpha_d v_d - v_{d+1}) - (\alpha - 1)(\beta_1 v_1 + \dots + \beta_d v_d - v_{d+2}) = 0$$

which demonstrates affine dependence, as the sum of coefficients is

$$(\beta - 1)(\alpha - 1) - (\alpha - 1)(\beta - 1) = 0.$$

Convex: A set $C \subset \mathbb{R}^d$ is *convex* if for any two points $v_1, v_2 \in C$, the interval joining these two points lies in C . That is, if $v_1, v_2 \in C$, then $\{\alpha v_1 + (1 - \alpha)v_2 : 0 \leq \alpha \leq 1\} \subset C$.

Convex hull: For $X \subset \mathbb{R}^d$, the *convex hull* $\text{conv}(X)$ is the intersection of all convex sets containing X . Then

$$\text{conv}(X) = \{\alpha_1 v_1 + \dots + \alpha_n v_n : n \in \mathbb{N}, v_1, \dots, v_n \in X, \sum_i \alpha_i = 1, \alpha_i \geq 0\}$$

Note that $\alpha_1 v_1 + \dots + \alpha_n v_n$ where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$ is called a *convex combination* of v_1, \dots, v_n .

To prove this formula, we note by induction on n we see that any convex combination of n vectors in X is in $\text{conv}(X)$: wlog $\alpha_n \neq 0$, $\alpha_n \neq 1$. Then

$$\alpha_1 v_1 + \dots + \alpha_n v_n = (1 - \alpha_n) \underbrace{\left[\frac{\alpha_1}{1 - \alpha_n} v_1 + \dots + \frac{\alpha_{n-1}}{1 - \alpha_n} v_{n-1} \right]}_{\in \text{conv}(X)} + \alpha_n \underbrace{v_n}_{\in \text{conv}(X)}$$

So this set of convex combinations is a subset of $\text{conv}(X)$. But since the set of convex combinations is itself convex, as it is easy to check, we have equality.

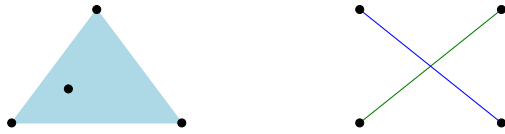
Convex dependence: We say $X \subset \mathbb{R}^d$ is *convex dependent* if for some $x \in X$ we have $x \in \text{conv}(X - \{x\})$.

We say X is *convex independent* if it is not convex dependent.

Remark. No point on the boundary of a circle is a convex combination of the other points, so there is no bound on the size of convex independent sets.

Theorem 8.1 (Radon 1921). *Let $A \subset \mathbb{R}^d$ be finite, with $|A| \geq d + 2$. Then there exist disjoint $A_1, A_2 \subset A$ such that $\text{conv}(A_1) \cap \text{conv}(A_2) \neq \emptyset$.*

Remark. Let's first consider the case $d = 2$, $|A| = 4$. If three points do not fall on a line (a degenerate case), then there are two cases: In



That is, one vector in X is a convex combination of some of the others.

Figure 41: Four points in \mathbb{R}^2 , when no three form a line. Either one is in the convex hull of the other three (left), or we may divide into pairs so that the convex hulls intersect (right).

either case, we may verify Radon's theorem.

Proof. WLOG, we may take $|A| = d + 2$; any surplus points we could put in A_1, A_2 , or neither, without changing the result. Set $A = \{v_1, v_2, \dots, v_{d+2}\}$. Since v_1, v_2, \dots, v_{d+2} are affinely dependent[†], there

[†] since there are more than $d + 1$ of them

exist $\alpha_1, \dots, \alpha_{d+2}$ with $\sum_i \alpha_i = 0$ not all zero such that

$$\alpha_1 v_1 + \dots + \alpha_{d+2} v_{d+2} = 0.$$

Let $A_1 = \{v_i : \alpha_i > 0\}$ and $A_2 = \{v_i : \alpha_i < 0\}$. WLOG, $\alpha_1, \dots, \alpha_k > 0$ and $\alpha_{k+2}, \dots, \alpha_{d+2} < 0$. Let $s = \alpha_1 + \dots + \alpha_k = -\alpha_{k+1} - \dots - \alpha_{d+2}$. Then

$$\left(\frac{\alpha_1}{s}\right) v_1 + \dots + \left(\frac{\alpha_k}{s}\right) v_{k+1} = \left(-\frac{\alpha_{k+1}}{s}\right) v_{k+1} + \dots + \left(-\frac{\alpha_{d+2}}{s}\right) v_{d+2}.$$

But as shown by the LHS, this quantity lies in $\text{conv}(A_1)$, and as shown by the RHS, the quantity lies in $\text{conv}(A_2)$. \square

Theorem 8.2 (Carathéodory 1911). *Every point in $\text{conv}(X)$ for $X \subset \mathbb{R}^d$ is a convex combination of at most $d + 1$ points in X .*

Proof. Let $x \in \text{conv}(X)$ be written $x = \alpha_1 v_1 + \dots + \alpha_n v_n$ for $v_1, \dots, v_n \in X$ and $\sum_i \alpha_i = 1$ with $\alpha_i \geq 0$ not all zero. Suppose n is chosen minimally such that a convex combination exists. Suppose $n \geq d + 2$ for the sake of contradiction. Then by Radon's theorem, wlog

$$\text{conv}(\{v_1, \dots, v_k\}) \cap \text{conv}(\{v_{k+1}, \dots, v_n\}).$$

So, for some $\beta_1, \dots, \beta_n \geq 0$,

$$\beta_1 v_1 + \dots + \beta_k v_k = \beta_{k+1} v_{k+1} + \dots + \beta_n v_n$$

with $\sum_{i=1}^k \beta_i = \sum_{i=k+1}^n \beta_i = 1$. Then

$$x = (\alpha_1 + \epsilon \beta_1) v_1 + \dots + (\alpha_k + \epsilon \beta_k) v_k + (\alpha_{k+1} - \epsilon \beta_{k+1}) v_{k+1} + \dots + (\alpha_n - \epsilon \beta_n) v_n \quad (\star)$$

has the sum of coefficients one for every ϵ . If $\epsilon > 0$ is minimally such that $\alpha_i - \epsilon \beta_i = 0$ for some i , then the expression (\star) is a convex combination of $\{v_1, \dots, v_n\} - \{v_i\}$, giving a contradiction. \square

Theorem 8.3 (Helly 1923). *Let C_1, C_2, \dots, C_n be a collection of convex sets in \mathbb{R}^d . If $\bigcap_{i \in I} C_i \neq \emptyset$ for every $I \subset [n]$ with $|I| = d + 1$, then $\bigcap_{i=1}^n C_i \neq \emptyset$.*

Remark. For \mathbb{R}^1 , if we consider $C_i = [a_i, b_i]$, then if each $C_i \cap C_j \neq \emptyset$, then we need $a_i \leq b_j$ for each i, j . Then $[\max_i a_i, \min_i b_i] \subset C_k$ for all k .

Convex sets in \mathbb{R}^1 may be open or closed intervals or rays, but we will take closed intervals for simplicity.

Proof by induction on n . Let us postpone the base case, $n = d + 2$. For the induction step, let $C'_1 = C_1 \cap C_n, \dots, C'_{n-1} = C_{n-1} \cap C_n$. Then by the base case, any $(d + 1)$ sets in $\{C'_1, \dots, C'_{n-1}\}$ have a non-empty intersection. Therefore they all intersect by the induction hypothesis, hence

$$\bigcap_{i=1}^n C_i = \bigcap_{i=1}^{n-1} C'_i \neq \emptyset.$$

So let us show the base case. Let $v_i \in \bigcap_{j \neq i} C_j$; by the assumption. Let $A = \{v_1, \dots, v_{d+2}\}$. Then by Radon's theorem, there exist disjoint $A_1, A_2 \subset A$ such that $\text{conv}(A_1) \cap \text{conv}(A_2) \neq \emptyset$. Consider $v \in \text{conv}(A_1) \cap \text{conv}(A_2)$. We will show that $v \in C_i$ for every i .

Suppose wlog that $v_i \in A_2$. Then $A_1 \subset C_i$. Since C_i is convex, then $\text{conv}(A_1) \subset C_i$. So $v \in C_i$, as desired. See fig. 42 for an example of this process in \mathbb{R}^2 . \square

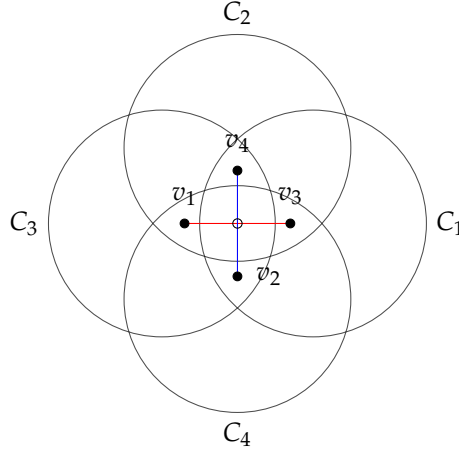


Figure 42: Helly's theorem in \mathbb{R}^2 . Given four convex sets C_1, \dots, C_4 such that every triple has a non-empty intersection $v_i \in \bigcap_{j \neq i} C_j$, we can use Radon's theorem to divide $\{v_1, v_2, v_3, v_4\}$ into two disjoint sets with non-empty convex hull. Here, $A_1 = \{v_1, v_3\}$ and $A_2 = \{v_2, v_4\}$. The point in the intersection $\text{conv}(A_1) \cap \text{conv}(A_2)$ is labelled here with an open circle.

Remark. Helly's theorem may not apply to infinite collections of sets; consider $\{[n, +\infty) : n \in \mathbb{N}\}$.

For any set of n points on the line, we can find a point such that there are at least $n/2$ points above it and below it. This is the median. How do we find an analog for \mathbb{R}^d ? We'd like to find a point that in any direction away from this point, there are still many points in our set.

Let us generalize to \mathbb{R}^d : Given $X \subset \mathbb{R}^d$, $|X| = n$, the *centerpoint* of x is a point $m \in \mathbb{R}^d$ such that for every closed halfspace $H \subset \mathbb{R}^d$ such that $m \in H$, then H contains at least $\frac{n}{d+1}$ points in X .

Theorem 8.4. *For every finite set $X \subset \mathbb{R}^d$, there exists a centerpoint.*

Proof. Note the following are all equivalent:

- m is a centerpoint
- For every closed halfspace $H \subset \mathbb{R}^d$, if $m \in H$ then $|H \cap X| \geq \frac{n}{d+1}$.
- For every closed halfspace $H \subset \mathbb{R}^d$, if $|X \cap H| < \frac{n}{d+1}$, then $m \notin H$.
- For every closed halfspace $H \subset \mathbb{R}^d$, if $|X \cap H^c| > n - \frac{n}{d+1} = \frac{dn}{d+1}$, then $m \in H^c$.
- For every open halfspace $H \subset \mathbb{R}^d$, if $|X \cap H| > \frac{dn}{d+1}$, then $m \in H$.

Consider the family

$$\mathcal{H} = \{H : H \text{ is an open halfspace of } \mathbb{R}^d, |H \cap X| > \frac{dn}{d+1}\}.$$

We will show for some $m \in \mathbb{R}^d$, we have $m \in H$ for every $H \in \mathcal{H}$, proving that m is a centerpoint. First, if $H_1, H_2, \dots, H_{d+1} \in \mathcal{H}$, then

$$H_1 \cap H_2 \cap \dots \cap H_{d+1} \neq \emptyset.$$

Otherwise, every point of X belongs to $\leq d$ out of $d+1$ of these half spaces, and thus

$$dn = \sum_{i=1}^{d+1} \frac{dn}{d+1} < \sum_{i=1}^{d+1} |H_i \cap X| \leq dn$$

which is a contradiction.

We would like to now apply Helly's theorem, but we have an infinite collection \mathcal{H} instead of a finite one.

Let us instead set $\mathcal{H}' = \{\text{conv}(Y) : Y \subset X, |Y| > \frac{dn}{d+1}\}$. By the previous argument and Helly's theorem, there exists $m \in \mathbb{R}^d$ such that $m \in C$ for every $C \in \mathcal{H}'$. But this suffices. For any $H \in \mathcal{H}$ there is $Y \subset X$ with $|Y| > \frac{dn}{d+1}$ and $Y \subset H$. Then $\text{conv}(Y) \subset H$ since H is convex. But since $m \in \text{conv}(Y)$, we have $m \in H$, as desired. \square

Theorem 8.5 (Birch). *Let $X \subset \mathbb{R}^2$ be a collection of $3n$ points. Then X can be partitioned into n triples such that the corresponding triangles all share a point in common.*

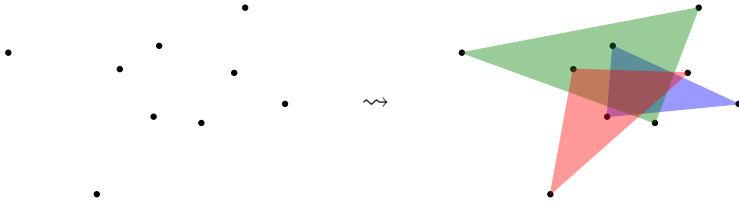


Figure 43: An example of Birch's theorem with nine points. On the left, the nine points are depicted; on the right, intersecting triangles chosen.

Remark. See fig. 43 for an example.

Proof. Let m be a centerpoint of X . Label the points of X by x_1, x_2, \dots, x_{3n} so that the rays $mx_1, mx_2, \dots, mx_{3n}$ are arranged around m in clockwise order, as shown in fig. 44. Let $X_i = \{x_i, x_{i+n}, x_{i+2n}\}$ for $i = 1, \dots, n$. Then $m \in \text{conv}(X_i)$ for each i .

Assume $m \notin \text{conv}(X_i)$. Then[†] some closed halfplane containing m separates m from the three points of X_i ; see fig. 45. But then the $2n$ rays between x_i and x_{i+2n} fall in the other half of the halfplane, so there are at least $2n+1$ points in X on the other side, contradicting that m is a centerpoint. \square

[†] We are appealing to a separating hyperplane theorem, such as Matoušek 2002, Theorem 1.24, although the result seems clear geometrically in this case.

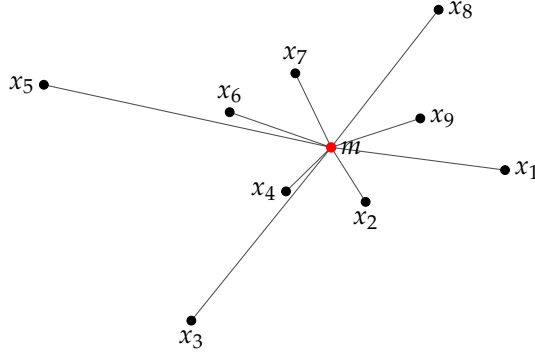


Figure 44: Nine points $\{x_1, \dots, x_9\}$ in \mathbb{R}^2 and their centerpoint m , in red. Continuing the example from fig. 43, the points are labelled clockwise in order from the centerpoint.

Theorem 8.6 (Colorful Carathéodory theorem, Bárány 1982). *Let $S_1, S_2, \dots, S_{d+1} \subset \mathbb{R}^d$ and suppose that $x \in \bigcap_{i=1}^{d+1} \text{conv}(S_i)$. Then there exist $x_1 \in S_1, x_2 \in S_2, \dots, x_{d+1} \in S_{d+1}$ such that $x \in \text{conv}(\{x_1, x_2, \dots, x_{d+1}\})$.*

Remark. The different sets correspond to different colors. Then x belongs to a convex combination of points each a different color. Taking each set $S_i \equiv X$, we recover Carathéodory's theorem.

Proof. We will say C is a *colorful simplex* if $C = \text{conv}(\{x_1, x_2, \dots, x_{d+1}\})$ for $x_i \in S_i$ for $i \in [d+1]$. Suppose that no colorful simplex contains x .

Choose a colorful simplex C such that $\text{dist}(C, x) = \min_{c \in C} \|c - x\|$ is minimal.

Let $z \in C$ be the closest point to x : $\text{dist}(z, x) = \text{dist}(C, x)$.

Let H be a hyperplane orthogonal to zx through z . Let H^+ and H^- be closed halfspaces with respect to H such that $x \in H^-$.

Claim. $C \subset H^+$.

Proof. Suppose not: $z' \in C \cap (H^- - H^+)$. Then angle $z'zx$ is acute and points on zz' near z are closer to x than z , a contradiction. ■

Claim. $z \in \text{conv}(\{x_1, \dots, x_{d+1}\} \cap H)$.

Proof. $z = \sum_{i=1}^{d+1} \lambda_i x_i$ with $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, since $z \in C$. Let $f: \mathbb{R}^d \rightarrow \mathbb{R}$ linear such that if $f(p) > 0$ for $p \in H^+ - H$, and $f(p) = 0$ for $p \in H$.

Then $0 = f(z) = \sum_{x_i \in H^+ - H} \lambda_i \underbrace{f(x_i)}_{>0} + \sum_{x_i \in H} \lambda_i \underbrace{f(x_i)}_{=0}$. So we must have $\lambda_i = 0$ when $x_i \in H^+ - H$. ■

By Caratheordy theorem applied to z and $\{x_i, \dots, x_{d+1}\} \cap H$, there exists j such that

$$z \in \text{conv}((\{x_1, x_2, \dots, x_{d+1}\} - \{j\}) \cap H).$$

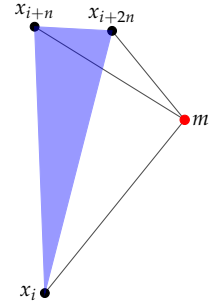


Figure 45: The case where the centerpoint m , in red, is not in the convex hull of X_i , shown in blue. Then the $2n$ points $\{x_i, x_{i+1}, \dots, x_{i+2n}\}$ all lie between x_i and x_{i+2n} (angularly) and may be separated from m by a closed hyperplane, contradicting the definition of centerpoint. $\|c - x\| = \text{dist}(c, x)$ is the 2-norm (Euclidean norm).

So $z' \notin H^+$

We know $x \in \text{conv}(S_j)$, so there exists $x'_j \in S_j \cap (H^- - H^+)$. Let $C' = \text{conv}(\{x_1, \dots, x_{j-1}, x'_j, x_{j+1}, \dots, x_{d+1}\})$.

Claim. Then $\text{dist}(C', x) < \text{dist}(C, x)$.

Proof. We know $z \in C'$. Then, as before, points on $x'_j z \subset C'$ are closer to x than z . ■

This last claim yields a contradiction to minimality of C .

Note: we assumed the S_i were finite when assuming there exists a minimal colorful simplex C . But in fact, we only need to consider $\text{conv}(S_i)$, which only depends on $d + 1$ points, by (the original) Carathéodory points. Thus, we could take $|S_i| = d + 1$ for each i . □

Remark. This proof yields an algorithm to improve our convex set one point at a time to the optimal one. Yet, it is unknown if this algorithm or any other can find the optimal $\text{conv}(\{x_1, \dots, x_{d+1}\})$ in polynomial time.

Theorem 8.7 (Tverberg 1966). *Let $A \subset \mathbb{R}^d$ with $|A| \geq (r - 1)(d + 1) + 1$. Then there exist $A_1, A_2, \dots, A_r \subset A$ pairwise disjoint such that $\bigcap_{i=1}^r \text{conv}(A_i) \neq \emptyset$.*

Remark. Radon's theorem is the case $r = 2$. If $d = 1$, then we have $2r - 1$ points in \mathbb{R} which we wish to write as r groups with intersecting convex hulls.

For $d = 2$, $3r - 2$ points in \mathbb{R}^2 can be partitioned into r groups with intersecting convex hulls. This in fact implies Birch's theorem.

If $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$, then $u \otimes v \in \mathbb{R}^{nm}$. Suppose $u = (u_1, \dots, u_n)$ and $v = (v_1, \dots, v_m)$; i.e., we have chosen bases for \mathbb{R}^n and \mathbb{R}^m . Then $u \otimes v$ can be thought of as a $m \times n$ matrix with components $(u \otimes v)_{ij} = u_i v_j$.

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad u \otimes v = \begin{pmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \end{pmatrix}.$$

Proposition.

1. $(\alpha_1 u_1 + \alpha_2 u_2) \otimes v = \alpha_1 (u_1 \otimes v) + \alpha_2 (u_2 \otimes v)$
2. Suppose v_1, \dots, v_k are linearly independent and $u_1 \otimes v_1 + u_2 \otimes v_2 + \dots + u_k \otimes v_k = 0$. Then $u_1 = u_2 = \dots = u_k = 0$.

We will leave the proof of the proposition as an exercise, and proceed to the proof of Tverberg's theorem.

Proof (Sarkaria 1992). Let $m = (r - 1)(d + 1) + 1$.

1. Instead of the original setting, consider $A = \{v_1, v_2, \dots, v_m\} \subset \mathbb{R}^{d+1}$ such that the vectors of A lie in an affine hyperplane not passing through the origin. I.e., there exists $f : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ linear such that $f(v_i) = 1$ for every i .
2. Let w_1, w_2, \dots, w_r be vectors in \mathbb{R}^{r-1} such that $w_1 + w_2 + \dots + w_r = 0$ and this is essentially the only relation between these vectors. For instance, we could choose $w_i = (\underbrace{0, \dots, 0}_{i-1}, 1, 0, \dots, 0)$ for $i = 1, \dots, r-1$ and $w_r = (-1, -1, \dots, -1)$.
3. Consider $\{v_i \otimes w_j : 1 \leq i \leq m, 1 \leq j \leq r\} \subset \mathbb{R}^{d+1} \otimes \mathbb{R}^{r-1} \cong \mathbb{R}^{m-1}$. For each i , we may think of $S_i := \{v_i \otimes w_j : 1 \leq j \leq r\}$ as a copy of the w_j embedded in a hyperplane. Note $0 \in \text{conv}(S_i)$ for each i , since

$$0 = \frac{1}{r} v_i \otimes (w_1 + w_2 + \dots + w_r) = \frac{1}{r} v_i \otimes w_1 + \frac{1}{r} v_i \otimes w_2 + \dots + \frac{1}{r} v_i \otimes w_r.$$
4. Apply the Colorful Carathéodory theorem to S_1, \dots, S_m and the point 0. We get that for $1 \leq i \leq m$ there exist j_i and $\lambda_i \geq 0$ such that

$$\sum_{i=1}^m \lambda_i (v_i \otimes w_{j_i}) = 0 \quad (2)$$

with $\sum_i \lambda_i = 1$.

Let $A_j = \{v_i : j_i = j\}$ for $j = 1, \dots, r$. Let us rewrite eq. (2) as follows.

$$\sum_{j=1}^r \left(\sum_{v_i \in A_j} \lambda_i v_i \right) \otimes w_j = 0.$$

Let us write $u_j = \sum_{v_i \in A_j} \lambda_i v_i$. Then we have

$$\begin{aligned} u_1 \otimes w_1 + u_2 \otimes w_2 + \dots + u_r \otimes w_r &= 0 \\ (u_1 - u_r) \otimes w_1 + (u_2 - u_r) \otimes w_2 + \dots + (u_{r-1} - u_r) \otimes w_{r-1} &= 0 \end{aligned}$$

using $w_r = -w_1 - w_2 - \dots - w_{r-1}$. But since $\{w_1, \dots, w_{r-1}\}$ are linearly independent, we must have $u_1 - u_r = u_2 - u_r = \dots = u_{r-1} - u_r = 0$. That is,

$$u_1 = u_2 = \dots = u_r.$$

Substituting the definition of u_j ,

$$\sum_{v_i \in A_1} \lambda_i v_i = \sum_{v_i \in A_2} \lambda_i v_i = \dots = \sum_{v_i \in A_r} \lambda_i v_i. \quad (3)$$

Suppose the sum of coefficients λ_i in each expression is the same and is equal to s . Then

In fact, $s = \frac{1}{r}$.

$$\sum_{v_i \in A_1} \frac{\lambda_i}{s} v_i = \cdots = \sum_{v_i \in A_r} \frac{\lambda_i}{s} v_i = p$$

and hence $p \in \text{conv}(A_k)$.

But if we apply f to eq. (3), by linearity we obtain

$$\sum_{v_i \in A_1} \lambda_i f(v_i) = \sum_{v_i \in A_2} \lambda_i f(v_i) = \cdots = \sum_{v_i \in A_r} \lambda_i f(v_i).$$

Since $f(v_i) = 1$ for all i , we find $\sum_{v_i \in A_j} \lambda_i \equiv s$. \square

Remark. In Helly's theorem, we consider $C_1, \dots, C_n \subset \mathbb{R}^d$ convex such that every $(d+1)$ -tuple of C 's has non-empty intersection. Then we obtain that there is a common point in all of the sets. Suppose instead $\sim \frac{1}{2}$ of the $(d+1)$ -tuples have a non-empty intersection. What can we say about large intersections?

Theorem 8.8 (Fractional Helly's theorem). *For every $d \in \mathbb{N}$ and $0 < \alpha \leq 1$ there exists a $\beta = \beta(\alpha, d)$ such that the following holds. Let $C_1, C_2, \dots, C_n \subset \mathbb{R}^d$ be convex and suppose $\bigcap_{i \in I} C_i \neq \emptyset$ for at least $\alpha \binom{n}{d+1}$ sets $I \subset [n]$ with $|I| = d+1$. Then there exists $X \subset [n]$ with $|X| \geq \beta n$ such that $\bigcap_{i \in X} C_i \neq \emptyset$.*

Proof. We need to assume that C_1, C_2, \dots, C_n are compact; we may do this as follows. For each set I as in the statment, select $p_I \in \bigcap_{i \in I} C_i$. Then replace C_i by $\text{conv}(\{p_I : I \text{ s.t. } I \ni i\})$.

Let $F_I = \bigcap_{i \in I} C_i$. Let $<$ be a linear lexicographic order on \mathbb{R}^d . Let $p_I = \min(F_I)$ in this order, if $F_I \neq \emptyset$. Since F_I is compact, p_I exists and is unique.

Claim. For every $I \subset [n]$ with $|I| = d+1$, s.t. $F_I \neq \emptyset$, there exists $J \subset I$, $|J| = d$ such that $p_I = p_J$.

Remark. For any $J \subset I$, we have $p_J \leq p_I$.

Proof of claim. Let $C = \{q \in \mathbb{R}^d : q < p_I\}$. Then C is convex. Then $C \cap (\bigcap_{i \in I} C_i) = C \cap F_I = \emptyset$. Since this $(d+2)$ -tuple intersection is empty, then by the contrapositive of Helly's theorem, not all $(d+1)$ -tuples can have empty intersection. But since $F_I \neq \emptyset$, the $(d+1)$ -tuple with empty intersection must include C . So there exists $J \subset I$ with $|J| = d$ such that

$$C \cap \left(\bigcap_{i \in J} C_i \right) = \emptyset.$$

Then $p \geq p_I$ for every $p \in F_J$. \blacksquare

Now for every $I \subset [n]$ with $|I| = d + 1$, $F_I \neq \emptyset$, select $J \subset I$ with $|J| = d$ such that $P_J = P_I$. Then there are $\alpha \binom{n}{d+1}$ sets I which we consider, and $\binom{n}{d}$ sets J of size d . So some set J_0 is associated with at least $\frac{\alpha \binom{n}{d+1}}{\binom{n}{d}}$ sets I . Such sets are of the form $J_0 \cup \{i\}$ for some i , and $p_{J_0} \in C_i$ for such i . So p_{J_0} belongs to at least

$$\frac{\alpha \binom{n}{d+1}}{\binom{n}{d}} + d = \frac{\alpha(n-d)}{d+1} + d \geq \frac{\alpha}{d+1}n$$

sets C_i (counting the d sets in J_0). Therefore, we may take $\beta = \frac{\alpha}{d+1}$. \square

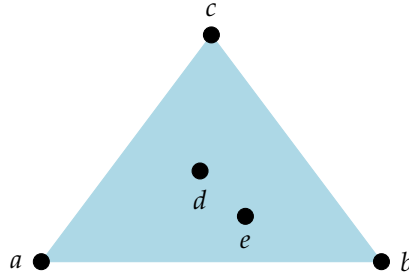
Remark. This constant is not optimal; for instance, when $\alpha = 1$, we would like $\beta = 1$ to recover Helly's theorem. The optimal constant is known, however.

For every k , we'd like to prove existence (and estimate the value) of the number $g(k)$ such that in every set of $g(k)$ points in \mathbb{R}^2 , with no three lying on a line[†] there exist k points which are convexly independent[‡].

Let us consider small k . Then $g(1) = 1$, $g(2) = 2$, $g(3) = 3$. But $g(4) > 4$, as shown in fig. 46.

Lemma 8.9. $g(4) = 5$.

Proof. Let $X \subset \mathbb{R}^2$ be a set of five points in general position. Consider $\text{conv}(X)$: this is a polygon with 3, 4, or 5 vertices. If there are 4 or 5 vertices we are done, so, assume $\text{conv}(X)$ is a triangle with vertices $\{a, b, c\}$. Let $\{d, e\}$ be the remaining two points of X , as shown in fig. 47.



Now assume the line de intersects ab and ac , as shown in fig. 48.

Then $bcd e$ form a convex quadrangle. We may see this by taking the convex hull of any three of $\{b, c, d, e\}$ to form a triangle T , and calling the omitted point z . Then there are two boundary line segments of the quadrilateral $bcd e$ which end in z , but each of these line segments only intersects T at the opposite end from z . If $z \in T$, then

[†] "in general position"

[‡] "in convex position," or are the vertices of a convex polygon

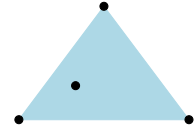


Figure 46: An illustration of the fact that $g(4) > 4$. We have four points with no three lying on a line such that not all four are convexly independent: one point is in the convex hull of the other three.

Figure 47: We assume $\text{conv}(X) = \text{conv}\{a, b, c\}$; then $\{d, e\} := X - \{a, b, c\}$ lie inside the triangle formed by $\{a, b, c\}$.

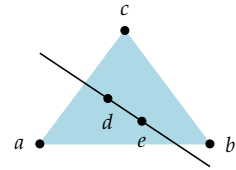


Figure 48: We assume the line passing through d and e intersects the sides ac and ab of the triangle. Since no three points in X are collinear, the line passing through d and e must intersect two of the sides of the triangle, so we can always relabel the vertices so that this is the case.

the whole line between z and those endpoints would have to be included in T . Thus, $z \notin T$, and we truly have a convex quadrilateral $bcd e$. This is illustrated in fig. 49. \square

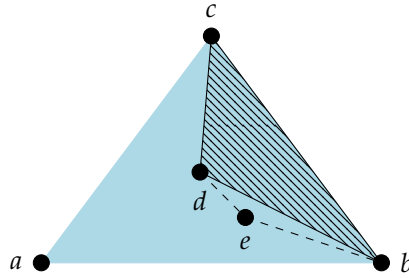


Figure 49: We will take $T = \text{conv}(\{c, d, b\})$, the crosshatched region. Then the two dashed lines only intersect T at the vertices d, b . We see if e were moved up into T , then the line passing through d and e would no longer intersect the segment ab .

Theorem 8.10 (Erdős and Szekeres 1935). $g(k)$ exists for all k .

Proof. Let $n = g(k)$ be such that in every coloring of $[n]^{(4)}$ in colors red and blue, one can find either a set of 5 with all quadruples in red, or a set of size k with all quadruples blue. This exists by the Hypergraph Ramsey theorem[†].

Color quadruples of points in red color if it is not in convex position, and otherwise in blue. By lemma 8.9, there exists a set of k points such that every four of them are in convex position.

Then these k points are in convex position by Carathéodory's theorem. \square

[†] For all positive integers r, k_1 , and k_2 there exists a positive integer $n = R^{(r)}(k_1, k_2)$ so that the following holds. If elements of $[n]^{(r)}$ are colored in colors red and blue then there is a set $Z \subset [n]$ such that either $|Z| = k_1$ and all elements of $Z^{(r)}$ are red, or $|Z| = k_2$ and all elements of $Z^{(r)}$ are blue.

Remark. We may obtain much better bounds by longer proofs.

References for Section 8.

- Bárány, I. (1982). "A generalization of Carathéodory's theorem". In: *Discrete Mathematics* 40.2, pp. 141–152 (cit. on p. 65).
- Carathéodory, C. (1911). "Über den variabilitätsbereich der fourier'schen konstanten von positiven harmonischen funktionen". In: *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 32.1, pp. 193–217. ISSN: 0009-725X. DOI: [10.1007/BF03014795](https://doi.org/10.1007/BF03014795) (cit. on p. 62).
- Helly, E. (1923). "Über Mengen konvexer Körper mit gemeinschaftlichen Punkte." In: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 32, pp. 175–176. URL: <http://eudml.org/doc/145659> (cit. on p. 62).
- Matoušek, J. (2002). *Lectures on discrete geometry*. Vol. 212. Graduate Texts in Mathematics. Springer. DOI: [10.1007/978-1-4613-0039-7](https://doi.org/10.1007/978-1-4613-0039-7) (cit. on p. 64).
- Erdős, P. and G. Szekeres (1935). "A combinatorial problem in geometry". In: *Compositio Mathematica* 2, pp. 463–470 (cit. on pp. 49, 57, 70).
- Radon, J. (1921). "Mengen konvexer Körper, die einen gemeinsamen Punkt enthalten". In: *Mathematische Annalen* 83.1, pp. 113–115. ISSN: 1432-1807. DOI: [10.1007/BF01464231](https://doi.org/10.1007/BF01464231) (cit. on p. 61).

- Sarkaria, K. S. (1992). "Tverberg's theorem via number fields". In: *Israel journal of mathematics* 79.2, pp. 317–320 (cit. on p. 66).
- Tverberg, H. (1966). "A Generalization of Radon's Theorem". In: *Journal of the London Mathematical Society* s1-41.1, pp. 123–128. DOI: [10.1112/jlms/s1-41.1.123](https://doi.org/10.1112/jlms/s1-41.1.123) (cit. on p. 66).

9 Incidence problems

Theorem 9.1 (Euler's formula). *Let G be a connected graph drawn in the plane without crossings. Then*

$$|V(G)| - |\mathcal{E}(G)| + \text{Reg}(G) = 2$$

where $\text{Reg}(G)$ is the number of regions in which the drawing divides the plane.

Corollary 9.2. *Let G be a graph drawn in the plane without crossings. Then*

$$|\mathcal{E}(G)| \leq 3|V(G)|.$$

Proof. We assume $|V(G)| \geq 3$. By adding additional edges we can ensure that each region of the drawing has a cycle of length 3 as its boundary. Then

$$3\text{Reg}(G) = 2|\mathcal{E}(G)|$$

by double counting pairs (edge, region) such that each edge belongs to two regions boundary. Substituting into Euler's formula, we have

$$|V(G)| - |\mathcal{E}(G)| + \frac{2}{3}|\mathcal{E}(G)| = 2$$

so

$$|\mathcal{E}(G)| = 3|V(G)| - 6 \leq 3|V(G)|.$$

□

Let $\text{cr}(G)$ denote the minimal number of pairs of crossing edges taken over all drawing of G in the plane where vertices are represented by points, edges by curves joining corresponding points, and the drawing of edges are allowed to intersect, and each intersection is a *crossing*[†].

Then $\text{cr}(G) = 0$ iff G can be drawn in the plane without crossings, and $\text{cr}(K_5) = 1$, by the drawing

[†] locally looks like an X , not two curves bouncing off each other.

Corollary 9.3. $\text{cr}(G) \geq |\mathcal{E}(G)| - 3|V(G)|.$

Proof. Let G be a graph with $\text{cr}(G) = c$. In a drawing of G with c pairs of crossing edges, remove c edges to obtain a graph drawn without crossings. Then by corollary 9.2,

$$|\mathcal{E}(G)| - c \leq 3|V(G)|$$

as we wanted. □

Theorem 9.4 (Crossing number lemma).

$$\text{cr}(G) \geq \frac{1}{64} \frac{m^3}{n^2}$$

for every graph G with m edges and n vertices such that $m \geq 4n$.

Proof. Let $p \in [0, 1]$ which we will choose later. Let $X \subset V(G)$ be obtained by choosing to include each vertex independently at random with probability p . Let G' be the random subgraph of G induced by X , namely $V(G') = X$, $\mathcal{E}(G') = \mathcal{E}(G|X)$. Let $m = |\mathcal{E}(G)|$, $n = |V(G)|$, $x = \text{cr}(G)$, $m' = |\mathcal{E}(G')|$, $n' = |V(G')|$, $x' = \text{cr}(G')$. Then by construction $\mathbb{E}[n'] = np$. Since each edge of G survives with probability p^2 , by linearity $\mathbb{E}[m'] = mp^2$. Next, $\mathbb{E}[x'] \leq xp^4$ because the probability that a particular pair of crossing edges survives is p^4 .

By corollary 9.3, $x' \geq m' - 3n'$. Taking expectation values, we have the bound $xp^4 \geq mp^2 - 3np$. That is,

$$x \geq \frac{m}{p^2} - 3 \frac{n}{p^3}.$$

Now, we choose optimal p :

$$\begin{aligned} \frac{\partial}{\partial p} \left(\frac{m}{p^2} - \frac{3n}{p^3} \right) &= -\frac{2m}{p^3} + \frac{9n}{p^4} = 0 \\ \implies p &= \frac{1}{4.5} \frac{n}{m}. \end{aligned}$$

Let us simply take $p = \frac{1}{4} \frac{n}{m}$. Then

$$\text{cr}(G) := x \geq \frac{m}{\frac{16n^2}{m^2}} - \frac{3n}{\frac{64n^3}{m^3}} p = \frac{m^3}{n^2} \left(\frac{1}{16} - \frac{3}{64} \right) = \frac{1}{64} \frac{m^3}{n^2}.$$

□

Let P be a set of points in the plane. Let L be a set of lines. Then define

$$I(P, L) := |\{(p, L) : p \in P, l \in L, p \in l\}|.$$

Then $I(P, L)$ is the number of *incidences* of P and L . Let $I(m, n)$ denote the maximum $I(P, L)$ over sets P, L with $|P| = m$, $|L| = n$.

Clearly, $I(m, n) \leq mn$.

Example. We see $I(3, 3) \leq 6$ by considering cases: if the three lines are parallel, the number of incidences is at most three. If two lines are parallel, there are at most five incidences. If none of the lines are parallel, there are at most six incidences. On the other hand, fig. 50 shows $I(3, 3) \geq 6$.

Example. Let $n = 4k^3$. Set

$$P = \{(x, y) : 0 \leq x \leq k-1, 0 \leq y \leq 4k^2-1, x, y \in \mathbb{Z}\}$$

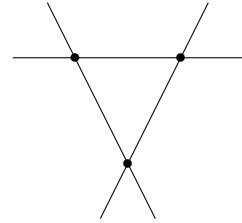


Figure 50: We see $I(3, 3) \geq 6$.

and

$$L = \{y = ax + b : 0 \leq a \leq 2k - 1, 0 \leq b \leq 2k^2 - 1, a, b \in \mathbb{Z}\}$$

Then $|P| = |L| = n$. We have $I(P, L) = k \cdot |L| = kn \geq \frac{1}{2}n^{4/3}$
every line in L incident with k points in P .

$$(x, ax + b) \in P$$

$$\text{for } 0 \leq x \leq k - 1, 0 \leq a \leq 2k - 1, 0 \leq b \leq 2k^2 - 1$$

$ax + b \leq (k - 1)(2k - 1) + 2k^2 - 1 < 4k^2$. So $I(n, n) \geq cn^{4/3}$ for all n
and some $c > 0$.

Theorem 9.5 (Szemerédi and Trotter 1983).

$$I(m, n) \leq 4m^{2/3}n^{2/3} + 4m + n.$$

Proof. Let P, L be sets of points and lines such that $|P| = m$, $|L| = n$
and $I(P, L) = I(n, m)$. Let G be a graph drawn in the plane with
crossings such that $V(G) = P$, and $\mathcal{E}(G)$ are drawn as line segments
joining consecutive points on lines in L . Then

$$\text{cr}(G) \leq n^2$$

I.e., follow one line at a time, connecting
consecutive points.

because every crossing in G is an intersection of two lines in L (loose
estimate). Then

$$I := I(P, L) = |\mathcal{E}(G)| + n$$

if there is a point of P on every line in L (if not, delete that line). We
see this by following each line in L , and noting between every two
incidences, we have an edge: if there are k points in P on a single
line, there are $k - 1$ edges connecting them.

Then $|\mathcal{E}(G)| = I - n$. We have $|\mathcal{E}(G)| \leq 4|V(G)|$ if $I - n \leq 4m$, that
is $I \leq 4m + n$.

Otherwise, by theorem 9.4,

$$n^2 \geq \text{cr}(G) \geq \frac{1}{64} \frac{|\mathcal{E}(G)|^3}{|V(G)|^2} = \frac{1}{64} \frac{(I - n)^3}{m^2}.$$

That is,

$$64m^2n^2 \geq (I - n)^3$$

$$4m^{2/3}n^{2/3} \geq I - n$$

$$I \leq 4m^{2/3}n^{2/3} + n. \quad \square$$

Conjecture (Erdős and Szemerédi 1983). For every $\epsilon > 0$ there exists
 $c_\epsilon > 0$ s.t. for every $A \subset \mathbb{Z}$,

$$|A + A| + |A \cdot A| \geq c_\epsilon |A|^{2-\epsilon},$$

where $A + A = \{a + b : a, b \in A\}$ and $A \cdot A = \{a \cdot b : a, b \in A\}$.

Theorem 9.6 (Elekes 1997). *There exists $c > 0$ such that for every $A \subset \mathbb{Z}$,*

$$|A + A| \cdot |A \cdot A| \geq c|A|^{5/2}.$$

In particular, by the arithmetic-geometric inequality,

$$|A + A| + |A \cdot A| \geq c|A|^{5/4}.$$

Proof. Let

$$P = \{(a, b) : a \in A + A, b \in A \cdot A\}.$$

Then $|P| = |A + A| \cdot |A \cdot A|$. Choose

$$L = \{y = a(x - b) : a, b \in A\}.$$

Then $|L| = |A|^2$. Next,

$$I(P, L) \geq |A|^3$$

since each line in L contains $|A|$ points in P as follows: $y = a(x - b)$, so choose $x = b + a' \in A + A$ for any $a' \in A$. Then $y = aa' \in A \cdot A$, so $(x, y) \in P$.

Therefore, by theorem 9.5,

$$|A|^3 \leq I(P, L) \leq 4|A|^{4/3}|A + A|^{2/3}|A \cdot A|^{2/3} + 4|A + A| \cdot |A \cdot A| + |A|^2.$$

At least one of these three terms is $\frac{1}{3}$ of the LHS. Easy if it is not the first one. Otherwise: $\frac{1}{3}|A|^3 \leq 4|A|^{4/3}|A + A|^{2/3}|A \cdot A|^{2/3}$, so

$$\left(\frac{1}{12}\right)^{3/2} |A|^{5/2} \leq |A + A| \cdot |A \cdot A|. \quad \square$$

References for Section 9.

- Elekes, G. (1997). “On the number of sums and products”. In: *Acta Arithmetica* 81.4, pp. 365–367 (cit. on pp. 4, 75).
- Erdős, P. and E. Szemerédi (1983). “Studies in Pure Mathematics: To the Memory of Paul Turán”. In: ed. by P. Erdős et al. Basel: Birkhäuser Basel. Chap. On sums and products of integers, pp. 213–218. ISBN: 978-3-0348-5438-2. DOI: [10.1007/978-3-0348-5438-2_19](https://doi.org/10.1007/978-3-0348-5438-2_19) (cit. on pp. 4, 74).
- Szemerédi, E. and W. T. Trotter (1983). “Extremal problems in discrete geometry”. In: *Combinatorica* 3.3, pp. 381–392. ISSN: 1439-6912. DOI: [10.1007/BF02579194](https://doi.org/10.1007/BF02579194) (cit. on pp. 4, 74).

10 Algebraic methods

10.1 Combinatorial Nullstellensatz (Alon and Tarsi 1989)

FOR BACKGROUND we will first consider the following result about zeros of multivariate polynomials.

Theorem (Hilbert Nullstellensatz (Hilbert 1893)). *Let \mathcal{F} be an algebraically closed field. Let $g_1, g_2, \dots, g_m \in \mathcal{F}[x_1, x_2, \dots, x_n]$ be polynomials of n variables over \mathcal{F} . Suppose $f \in \mathcal{F}[x_1, \dots, x_n]$ is such that $f = 0$ whenever $g_1 = g_2 = \dots = g_m = 0$. Then there exists $n \in \mathbb{N}$ and polynomials $h_1, \dots, h_m \in \mathcal{F}[x_1, \dots, x_n]$ such that $f^n = h_1 g_1 + h_2 g_2 + \dots + h_m g_m$.*

In one variable: $g = (x - a_1) \cdot (x - a_2) \cdots (x - a_z)$. Assume all roots are distinct. Then $f = 0$ at zeros of g iff $f = gh$ for some h . If not all roots are distinct, then $f = 0$ at zeros of g iff $f^k = gh$ for some $k \in \mathbb{N}$.

LET US PROCEED to Combinatorial Nullstellensatz. Recall the degree of a multivariable monomial is the sum of degrees in each variable (e.g. $\deg(x_1^2 x_2^3) = 5$), and the degree of a polynomial is as usual the maximum degree of its monomials.

Theorem 10.1. *Let \mathcal{F} be a field, $S_1, S_2, \dots, S_n \subset \mathcal{F}$,*

$$g_i = \prod_{s \in S_i} (x_i - s)$$

for $i = 1, 2, \dots, n$. If $f \in \mathcal{F}[x_1, \dots, x_n]$ such that f vanishes on all common zeros of g_1, g_2, \dots, g_n (i.e., $f(x_1, \dots, x_n) = 0$ if $x_i \in S_i$ for all i), then there exist polynomials $h_1, h_2, \dots, h_n \in \mathcal{F}[x_1, \dots, x_n]$ with $\deg h_i \leq \deg f - \deg g_i$ and $f = h_1 g_1 + h_2 g_2 + \dots + h_n g_n$.

Remark. The g_i are monic, single variable, with no double roots. This is very restrictive compared to Hilbert's Nullstellensatz, but we do obtain a stronger conclusion.

Given a multivariable polynomial $p(x_1, \dots, x_n)$, let us denote $\deg_{x_i} p$ as the “degree of p over x_i ”, i.e. the degree of p when considered as a polynomial in the variable x_i only, treating $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ as constants. To prove theorem 10.1, we'll need the following lemma.

The degree bound in theorem 10.1 means that each $h_i g_i$ has degree at most that of f .

Lemma 10.2. *Let $P \in \mathcal{F}[x_1, \dots, x_n]$ with $\deg_{x_i} P \leq t_i$ for $i = 1, \dots, n$. Let $S_1, \dots, S_n \subset \mathcal{F}$, $|S_i| \geq t_i + 1$ for $i \in [n]$. If $P(x_1, \dots, x_n) = 0$ when $x_i \in S_i$ for each $i \in [n]$, then $P \equiv 0$.*

Proof by induction on n . Base case ($n = 1$): a degree t polynomial has at most t roots. Induction step: we write

$$\begin{aligned} P(x_1, \dots, x_n) &= P_{t_n}(x_1, \dots, x_{n-1})x_n^{t_n} + \dots + P_1(x_1, \dots, x_{n-1})x_n + P_0(x_1, \dots, x_{n-1}) \\ &= \sum_{i=1}^{t_n} P_{t_i}(x_1, \dots, x_{n-1})x_n^i. \end{aligned}$$

for some polynomials P_{t_n}, \dots, P_0 in variables x_1, \dots, x_{n-1} . Now, fix $x_1 \in S_1, \dots, x_{n-1} \in S_{n-1}$. Then $P(x_1, \dots, x_n)$ becomes a one-variable polynomial in x_n of degree at most t_n with at least $t_n + 1$ roots (one for each $x_n \in S_n$). So by the base case, each $P_i(x_1, \dots, x_{n-1}) = 0$ for $i \in [t_n]$. Since we chose $x_1 \in S_1, \dots, x_{n-1} \in S_{n-1}$ arbitrarily, by the induction hypothesis, $P_i \equiv 0$. Hence $P \equiv 0$. \square

I.e., the P_{t_i} are the coefficients of x_n^i , and do not depend on x_n .

We may proceed to prove the theorem.

Proof of theorem 10.1. Let $t_i = |S_i| - 1$. For $i \in [n]$, we write

$$g_i = x_i^{t_i+1} - \sum_{j=0}^{t_i} a_{ij} x_i^j$$

for some coefficients a_{ij} . Given a polynomial f in variables x_1, \dots, x_n , we “divide with remainder” by each g_i , yielding

$$f = \left(\sum_{i=1}^n h_i g_i \right) + P.$$

with $\deg(h_i) \leq \deg(f) - \deg(g_i)$ and $\deg_{x_i}(P) \leq t_i$. Now, if f vanishes when all $x_i \in S_i$, then P does too, so lemma 10.2 we have $P \equiv 0$, giving the result. \square

We may reformulate this result into a perhaps more helpful form.

Theorem 10.3 (Combinatorial Nullstellensatz). *Let \mathcal{F} be a field, $f \in \mathcal{F}[x_1, \dots, x_n]$, $\deg(f) = t_1 + t_2 + \dots + t_n$ and the coefficient of $x_1^{t_1} x_2^{t_2} \dots x_n^{t_n}$ in f non-zero. Let $S_1, S_2, \dots, S_n \subset \mathcal{F}$ with $|S_i| \geq t_i + 1$. Then there exists $s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n$ such that $f(s_1, s_2, \dots, s_n) \neq 0$.*

Proof. Suppose not. Then f vanishes whenever $x_i \in S_i$, so by theorem 10.1, there exist polynomials h_1, \dots, h_n with $\deg(h_i) \leq \deg(f) - |S_i| \leq \deg(f) - t_i - 1$ such that

$$f = \sum_{i=1}^n h_i g_i$$

where $g_i = \prod_{s \in S_i} (x_i - s)$. Then the monomial $x_1^{t_1} x_2^{t_2} \dots x_n^{t_n}$ must appear with non-zero coefficient in some $h_i g_i$. Since $\deg f \geq \deg h_i + \deg g_i$, this term must be obtained by multiplying a monomial in h_i by the highest degree monomial in g_i , namely $x_i^{|S_i|}$. Since $|S_i| > t_i$, this is a contradiction. \square

Theorem (Cauchy-Davenport theorem). *Let $A, B \subset \mathbb{Z}/p\mathbb{Z} =: \mathbb{Z}_p$ for p prime. Then $|A + B| \geq \min\{p, |A| + |B| - 1\}$.*

Proof. Suppose $|A| + |B| \leq p + 1$ without loss of generality[†]. Suppose, for contradiction, that $|A + B| \leq |A| + |B| - 2$. Then select $C \supseteq A + B$ such that $|C| = |A| + |B| - 2$.

Let $a = |A|$ and $b = |B|$. Set $f(x, y) = \prod_{c \in C} (x + y - c)$. Then

$$\deg f = |C| = a + b - 2 = \overbrace{(a-1)}^{t_1} + \overbrace{(b-1)}^{t_2}$$

The coefficient of $x_1^{t_1} x_2^{t_2}$ in f is $\binom{a+b-2}{a-1}$. Note since $a + b - 2 \leq p + 1 - 2 = p - 1 < p$ so the binomial coefficient is not zero. Set $S_1 = A$, $S_2 = B$. By theorem 10.3, there exists $s_1 \in A$, $s_2 \in B$ such that $f(s_1, s_2) \neq 0$. Then $s_1 + s_2 \notin C$ which is a contradiction. \square

Let us consider a conjecture by Erdős and Heilbronn 1964, proven in 1994.

Theorem 10.4 (Da Silva and Hamidoune 1994). *Let $A, B \subset \mathbb{Z}_p$, with $A \oplus B = \{a + b : a \in A, b \in B, a \neq b\}$. Then*

$$|A \oplus B| \geq \min\{p, |A| + |B| - 2\}$$

if $|A| \neq |B|$.

Remark. The assumption $|A| \neq |B|$ is necessary. Say $A = \{0, 1, \dots, k\}$ with $k < p/2$, then $|A| = k + 1$ and

$$A \oplus A = \{1, 2, \dots, 2k - 1\}$$

so $|A \oplus A| = 2k - 1 = 2|A| - 3$.

Proof. Let $a = |A|$, $b = |B|$. WLOG, assume $a + b \leq p + 2$. Suppose for a contradiction that there exists $C \supset A \oplus B$ such that $|C| = a + b - 3$. Let

$$f(x, y) = (x - y) \prod_{c \in C} (x + y - c).$$

Then $\deg f = |C| + 1 = a + b - 2 = (a - 1) + (b - 1)$. Assume for now that the coefficient of $x^{a-1} y^{b-1}$ is non-zero. Then by theorem 10.3 there exists $s_1 \in A$, $s_2 \in B$ such that $f(s_1, s_2) \neq 0$. But then $s_1 \neq s_2$ and $s_1 + s_2 \notin C$. This is a contradiction.

It remains to find the coefficient of $x^{a-1} y^{b-1}$ in $f(x, y)$. The coefficient must be the same as in the polynomial $(x - y)(x + y)^{a+b-3}$. We use binomial expansion and only keep the terms of the right degree: $(x - y)(\alpha_1 x^{a-2} y^{b-1} + \alpha_2 x^{a-1} y^{b-2})$. Then the coefficient is

$$\alpha_1 - \alpha_2 = \binom{a+b-3}{a-2} - \binom{a+b-3}{a-1} \not\equiv 0 \pmod{p}$$

if and only if

$$\frac{(a+b-3)!}{(a-2)!(b-1)!} \not\equiv \frac{(a+b-3)!}{(a-1)!(b-2)!} \pmod{p}$$

which we may simplify to obtain

$$a - 1 \not\equiv b - 1 \pmod{p}.$$

[†] by throwing away elements of A and B . If the minimum m is achieved by p , we throw away elements until $|A| + |B| - 1 = p = m$. Then we show that the now-smaller $|A + B|$ has $|A + B| \geq m$. Then the original $|A + B|$ must be at least m as well.

\square

n-dimensional hypercube $Q_n := \{0,1\}^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \in \{0,1\}\}$. 2^n points. See fig. 51 for Q_3 . We'd like to cover Q_n by affine hyperplanes; this can be done with 2: one on top and one on bottom. Suppose we'd like to cover exactly all vertices except the origin.

Theorem 10.5 (Alon and Füredi 1993). *Let H_1, H_2, \dots, H_m be affine hyperplanes in \mathbb{R}^n with cover exactly $2^n - 1$ vertices of the *n*-dimensional hypercube Q_n . Then $m \geq n$.*

Proof. Let $H_i = \{x : \langle a_i, x \rangle = b_i\}$ for some $a_i = (a_{i1}, \dots, a_{in}) \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. That is, if $x = (x_1, \dots, x_n) \in H_i$,

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i.$$

Assume for contradiction that $m < n$, and H_1, \dots, H_m do not cover the vertex $(0,0,\dots,0)$, but cover every $x = (x_1, \dots, x_n)$ for $x_i \in \{0,1\}$ with not all $x_i = 0$. Define

$$h(x_1, \dots, x_n) = \prod_{i=1}^m (\langle a_i, x \rangle - b_i).$$

Then by assumption, $h = 0$ for every vertex of Q_n except 0. Using the language of theorem 10.3, set $S_i = \{0,1\}$, $t_i = 1$. We want to choose a polynomial f with $\deg(f) = n$, the coefficient of $x_1 \cdots x_n$ non-zero, and $f(x) = 0$ for every vertex of Q_n . Our polynomial h will not suffice because $\deg h = m < n$ and in particular the coefficient of $x_1 \cdots x_n$ is zero. Instead, set

$$f(x_1, \dots, x_n) := (-1)^{m+n+1} \prod_{i=1}^m b_i \prod_{i=1}^n (x_i - 1) + h(x_1, \dots, x_n).$$

Then $\deg f = n$ and the coefficient of $x_1 \cdots x_n$ is $(-1)^{m+n+1} \prod_{i=1}^m b_i \neq 0$. Theorem 10.3 then yields some $x_1 \in S_1, \dots, x_n \in S_n$ such that $f(x_1, \dots, x_n) \neq 0$. Since if $x_i \in \{0,1\}$ not all zero, the first term of f vanishes (because at least one $x_i = 1$), and h vanishes by assumption, we must have $x_i \equiv 0$. But then

$$0 \neq f(0, \dots, 0) = -(-1)^m \prod_{i=1}^m b_i + h(0, \dots, 0) = -(-1)^m \prod_{i=1}^m b_i + (-1)^m \prod_{i=1}^m b_i$$

which is a contradiction. \square

Lemma 10.6. *Let A be an $n \times n$ matrix over a field \mathbb{F} . Suppose $\text{per}(A) \neq 0$. Let $S_1, S_2, \dots, S_n \subset \mathcal{F}$, $|S_i| = 2$ for every i . Let $b \in \mathcal{F}^n$. Then there exist $x_1 \in S_1, x_2 \in S_2, \dots, x_n \in S_n$ such that if $x = (x_1, \dots, x_n)$, then $(Ax)_j \neq b_j$ for every $j = 1, \dots, n$.*

Proof. Let $f(x_1, \dots, x_n) = \prod_{i=1}^n (a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n - b_i)$. By theorem 10.3, we can find $x_i \in S_i$ such that $f(x_1, \dots, x_n) \neq 0$, as desired, if the coefficient of x_1, \dots, x_n in f is non-zero. But this coefficient is exactly the permanent of A . \square

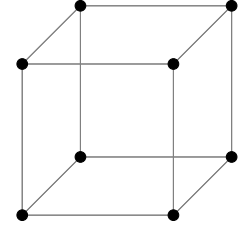


Figure 51: The cube Q_3 .

The vertex $\vec{0}$ was chosen without loss of generality, of course.

Note then $h(0, \dots, 0) = \prod_{i=1}^m (-b_i) \neq 0$.

The permanent of A , if $A = (a_{ij})$, is

$$\text{per}(A) = \sum_{\pi \in S_n} \prod_{i=1}^n a_{i\pi(i)},$$

i.e. is the determinant without the signs. For example, $\text{per} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad + bc$. Unlike the determinant, it is not known to be efficiently computable.

Let a_1, a_2, \dots, a_k be integers. We'd like to find nonempty $S \subset [k]$ such that $\sum_{i \in S} a_i$ is divisible by n . How large must k be (as a function of n)?

One needs at least n , as shown by taking $a_i = 1$ for every i .

In fact, $k = n$ suffices. Consider $\sigma_i = \sum_{j=1}^i a_j$ for $i = 1, \dots, n$, i.e., $\sigma_1 = a_1, \sigma_2 = a_1 + a_2$, etc. If $\sigma_i \not\equiv 0 \pmod{n}$ for every n , then there must exist i, j such that $\sigma_i \equiv \sigma_j \pmod{n}$. But then $\{i+1, \dots, j\}$ is the required set.

If $\sigma_i \equiv 0 \pmod{n}$ for some i , $\{1, 2, \dots, i\}$ suffices.

Now, let a_1, \dots, a_k be integers. We'd like to find $S \subset [k]$ with $|S| = m$ such that n divides $\sum_{i \in S} a_i$. If $m < n$, we can never guarantee the existence of S , as the example $a_i \equiv 1$ shows. In fact, we need at least m divisible by n . Let us take $m = n$.

Suppose $n = 2$. How many integers must we have such that two of them sum to an even number? $k = 3$; then there is always at least two evens or two odds.

Consider $n = 3$. We note $\{0, 0, 1, 1\}$ has no 3 numbers summing to a number divisible by 3. But $k = 5$ suffices.

In general, $k = 2n - 1$. For $2n - 2$, we may take $n - 1$ 0's and $n - 1$ 1's as a counterexample.

Theorem 10.7 (Erdős, Ginzburg, and Ziv 1961). *Let $a_1, a_2, \dots, a_{2n-1} \in \mathbb{Z}$ be integers. Then there exists $S \subset [2n - 1]$ with $|S| = n$ such that $\sum_{i \in S} a_i$ is divisible by n .*

Proof. Assume that $n = p$ is prime. We may reduce modulo p to assume $0 \leq a_i \leq p - 1$ and an ordering $a_1 \leq a_2 \leq \dots \leq a_{2p-1}$. Let $S_i = \{a_i, a_{i+p-1}\}$ for $i = 1, \dots, p - 1$. Let A be the all 1's matrix: $a_{ij} = 1$ for all $1 \leq i, j \leq p - 1$.

Let $\{b_1, \dots, b_{p-1}\}$ be all elements of \mathbb{Z}_p except $-a_{2p-1}$.

Example. $p = 3, a_1, a_2, \dots, a_5$. $S_1 = \{a_1, a_3\}, S_2 = \{a_2, a_4\}$. Suppose $a_5 = 2$. Then $\{b_1, b_2\} = \{0, 2\}$.

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \text{ or } a_3 \\ a_2 \text{ or } a_4 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

where the inequality in fact occurs componentwise.

Suppose there exists $s_1 \in S_1, \dots, s_{p-1} \in S_{p-1}$ such that for $s = (s_1, \dots, s_{p-1})$,

$$As \neq \begin{pmatrix} b_1 \\ \vdots \\ b_{p-1} \end{pmatrix}$$

where the inequality occurs for each coordinate. Equivalently, $s_1 + s_2 + \dots + s_{p-1} = -a_{2p-1} \pmod{p}$. That is, $s_1 + \dots + s_{p-1} + a_{2p-1} = 0 \pmod{p}$.

We may find such s_1, \dots, s_{p-1} using lemma 10.6, using that

$$\text{per} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix} = (p-1)! \neq 0 \pmod{p}.$$

However, we still need to check the condition $|S_i| = 2$ for each S_i . If for some i we have $|S_i| = 1$, then, $a_i = a_{i+p-1}$. In such a case, by our ordering, $a_i = a_{i+1} = \dots = a_{i+p-1}$, hence

$$\sum_{j=1}^{i+p-1} a_j = 0 \pmod{p}$$

as desired. The composite case will be left as an exercise by induction. \square

10.2 Kakeya needle problem (Kakeya 1917)

What is the minimum area of a region in the plane in which one can continuously rotate by 360° a needle of length 1? Figure 52 shows two attempts. In fact, the answer is zero.

Let us try to restate this in terms of finite fields. First, we have the equivalent[†] question: What is the minimum area of a region in the plane in which contains a segment of length 1 in every direction?

Now, we have the finite field Kakeya problem: What is the minimum $|E|$ such that $E \subset \mathbb{F}^d$ and E contains a line in every direction. That is, for every $v \in \mathbb{F}^d - \{0\}$ there exists x_v such that $x_v + tv \in E$ for every $t \in \mathbb{F}$. In particular, does there exist $c_d > 0$ (depending on d but not \mathbb{F}) such that

$$|E| \geq c_d |\mathbb{F}|^d$$

for all finite fields \mathbb{F} ?

Lemma 10.8. *Let $E \subset \mathcal{F}^d$ such that $|E| < \binom{k+d}{k}$. Then there exists non-zero polynomial f of d variables such $f(x) = 0$ for every $x \in E$, and $\deg f \leq k$.*

Proof. Let V be the vector space of polynomials in d variables over \mathcal{F} of degree at most k . Then $\dim V$ is the number of monomials $x_1^{k_1} x_2^{k_2} \dots x_d^{k_d}$ such that $k_1 + k_2 + \dots + k_d \leq k$. This is the same as the number of decompositions $k_1 + k_2 + \dots + k_d + k_{d+1} = k$ such that $k_i \geq 0, k_i \in \mathbb{Z}$. This is $\binom{k+d}{k}$ as follows: if we write k stars

★ ★ ★ ★ ★ ⋯ ★

then our task is to put d bars between the stars and thus partition the elements into k_1, k_2, \dots, k_{d+1}

★ | ★ ★ ★ | ★ ⋯ ★
 $k_1 \quad k_2$

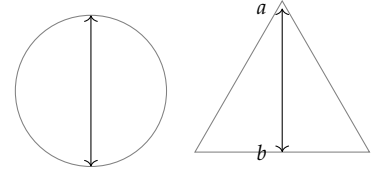


Figure 52: Left: we may simply rotate our needle of length about its middle, using a circle area of $\pi/4$. Right: we could rotate our needle by pulling end b to the bottom left corner of the triangle, then end a down the right side of the triangle. Then we could pull b up the left side of the triangle, and finally a to the middle, thereby completing the rotation. This uses an equilateral triangle of altitude 1, which then has area $\frac{1}{\sqrt{3}}$.

[†] It turns out these are equivalent, by adding regions of area zero connecting the segments in each direction.

so we must choose k stars from the $k + d$ total number of symbols.

Let $E = \{x_1, \dots, x_m\} \subset \mathbb{F}^d$. For $p \in V$, the map

$$p \mapsto (p(x^1), p(x^2), \dots, p(x^m))$$

is a linear map from V to $|E|$ -dimensional space. If $\dim(V) > |E|$, then there exists $p \in V$, $p \neq 0$ such that p is mapped to zero, i.e. p is zero everywhere on E . \square

Now, given \mathbb{F} a finite field, we will say $E \subset \mathbb{F}^d$ is a *Kakeya set* if for every $v \in \mathbb{F}^d - \{0\}$ there exists $x \in \mathbb{F}^d$ such that $x + tv \in E$ for every t .

Lemma 10.9. *If $E \subset \mathbb{F}^d$ is Kakeya, then there exists no $p \in \mathbb{F}[x_1, \dots, x_d]$, $p \neq 0$, $\deg p \leq |F| - 1$ such that $p(x) = 0$ for every $x \in E$.*

Proof. Suppose such p exists. Then

$$p(x) = p_k(x) + p_{k-1}(x) + \dots + p_0(x)$$

where $p_i(x)$ is homogenous polynomial of degree i and $p_k(x)$ not identically 0 ($\deg p = k$). By definition, for every $v \in \mathbb{F}^d$ there exists x such that $p(x + tv) = 0$ for every $t \in \mathbb{F}$. Fix x and v , $p(x + tv)$ is a polynomial in t of degree $\leq |F| - 1$. If $x = (x_1, \dots, x_d)$ and $v = (v_1, \dots, v_d)$, then $x + tv = (x_1 + tv_1, \dots, x_d + tv_d)$. The coefficient of t^k in p is $p_k(v)$.

On the other hand, p has $|F|$ roots so p is identically zero as a polynomial in t . In particular, $p_k(v) = 0$. But this is true for every v , contradicting the assumption that $p_k(v)$ is not identically zero. \square

Corollary 10.10 (Dvir 2009). *If $E \subset \mathbb{F}^d$ is Kakeya, then*

$$|E| \geq \binom{|F| - 1 + d}{d} = \frac{|\mathbb{F}|^d}{d!} + o(|\mathbb{F}|^d).$$

10.3 Shannon capacity (Shannon 1956)

We are transmitting messages in alphabet V over a noisy channel. Some symbols can be confused with each other during transmission.

Let G be a graph with vertex set V with a pair of symbols joined by an edge if they can be confounded.

Example. $V = \{a, b, d, p, q\}$.

We'd like to send a set S_n of n letter messages such that no two messages in S_n can be confused with each other. One solution: we can send 2^n messages which can't be confused just by using symbols a and p .

A *strong product* of graphs G and H denoted $G \boxtimes H$ is a graph with vertex set $V(G) \times V(H)$ and $(u_1, v_1) \sim (u_2, v_2)$ if

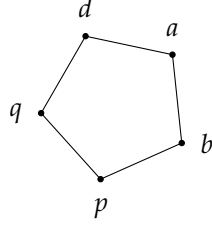


Figure 53: A graph representing transmitting the alphabet $V = \{a, b, d, p, q\}$. Adjacency indicates symbols which could be confused for each other. This graph is C_5 , the cycle of length five.

$u_1 = u_2$ or u_1 is adjacent to u_2

and $v_1 = v_2$ or v_1 is adjacent to v_2 .

Let $\boxtimes^n G$ denote $\underbrace{G \boxtimes G \boxtimes \cdots \boxtimes G}_{n \text{ times}}$.

We are interested in the maximum size of an independent set in $\boxtimes^n G$, which we will call $\alpha(\boxtimes^n G)$. Let $\Theta(G) = \lim_{n \rightarrow \infty} \sqrt[n]{\alpha(\boxtimes^n G)}$ be the *Shannon capacity*. This exists because $\alpha(G \boxtimes H) \geq \alpha(G)\alpha(H)$, as the product of two independent sets produces an independent set in the product graph.

Lemma 10.11. *For any graph, $\alpha(G) \leq \Theta(G) \leq \chi(\bar{G})$.*

For brevity, let us write $G^n := \boxtimes^n G$.

Proof. By the product inequality for α , we have $\alpha(G^n) \geq (\alpha(G))^n$, hence $\alpha(G) \leq \Theta(G)$.

Now, note $\alpha(G) \leq \chi(\bar{G})$, since $\alpha(G)$ is the size of the largest complete subgraph of \bar{G} . Further, $\chi(\bar{G} \boxtimes \bar{H}) \leq \chi(\bar{G})\chi(\bar{H})$.

We may think of coloring \bar{G} as coloring G such that any two vertices which are the same color are connected. To see $\chi(\bar{G} \boxtimes \bar{H}) \leq \chi(\bar{G})\chi(\bar{H})$, color each vertex (u, v) of $\bar{G} \boxtimes \bar{H}$ by the pair of color $(c_1(u), c_2(v))$ where c_1 is a coloring of \bar{G} and c_2 as a coloring of \bar{H} .

Now every color class is used on $K_l \boxtimes K_s$ for some l and s , but the strong product of complete graphs is complete. Therefore

$$\alpha(G^n) \leq \chi(\bar{G}^n) \leq (\chi(\bar{G}))^n$$

and we have the result as desired. \square

In fact, $\alpha(G) = \chi(\bar{G})$ for many graphs, including perfect graphs (hence all graphs on 4 vertices). The smallest graph in which $\alpha(G) \neq \chi(\bar{G})$ is the cycle of length five, C_5 , considered earlier. There, we saw we could achieve 2, and $\chi(\bar{C}_5) = 3$, so $2 \leq \Theta(C_5) \leq 3$.

We also have $\alpha(C_5^2) \geq 5$, that is, you can send 5 different 2 symbol messages. These messages are: $\{(1, 1), (2, 3)(3, 5), (4, 2), (5, 4)\}$, using the labelling of fig. 54. In particular, $2 < \sqrt{5} \leq \Theta(C_5)$, so our naive bound of 2 is not optimal.

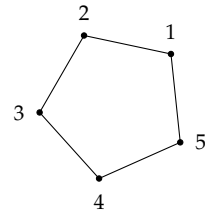


Figure 54: C_5 , the cycle of length 5.

To prove an upper bound on Θ , we'd like a function $\theta(G)$ such that

1. $\alpha(G) \leq \theta(G)$
2. $\theta(G \boxtimes H) \leq \theta(G)\theta(H)$.

Then, as in lemma 10.11, any such function will upper bound Θ , i.e., $\Theta(G) \leq \theta(G)$. Ideally, $\theta(C_5) = \sqrt{5}$, to prove the lower bound of $\sqrt{5}$ is optimal. This would confirm that we've found a sharper upper bound than $\chi(\bar{G})$.

One function satisfying all of these is the *Lovász theta*[†], defined as follows. Let G be a graph with $V(G) = [n]$. A collection $v_1, v_2, \dots, v_n \in \mathbb{R}^d$ is an *orthonormal representation* of G if

[†] Lovász 1979b

- $\|v_i\| = 1$,
- $v_i \perp v_j$ if i and j are not adjacent.

Let the *value* of an orthonormal representation be

Every graph has an orthonormal representation in dimension n .

$$\min_{\|c\|=1} \max_{v_i} \frac{1}{\langle v_i, c \rangle^2}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product (dot product). The vector c achieving the minimum is called a *handle*. Finally, define $\theta(G)$ as the minimal value over orthonormal representations.

That $\theta(G)$ satisfies points 1 and 2 is the content of lemmas 10.12 and 10.13. That $\theta(C_5) \leq \sqrt{5}$ is the content of lemma 10.14. This will conclude our discussion of Shannon capacity.

Lemma 10.12. $\alpha(G) \leq \theta(G)$

Proof. Let v_1, \dots, v_n be the optimal orthonormal representation of G with handle c . Let $\alpha(G) = k$, and suppose $\{1, 2, \dots, k\}$ form an independent set. Then v_1, \dots, v_k can be completed to an orthonormal basis, so

$$1 = \|c\|^2 = \langle c, c \rangle \geq \sum_{i=1}^k (\langle c, v_i \rangle)^2 \geq \frac{k}{\theta(G)}$$

since $\theta(G) \geq \frac{1}{\langle c, v_i \rangle^2}$. □

Lemma 10.13.

$$\theta(G \boxtimes H) \leq \theta(G)\theta(H).$$

Proof. Observation: if $u, v \in U$ and $x, y \in V$, then

$$\langle u \otimes x, v \otimes y \rangle = \langle u, v \rangle \langle x, y \rangle$$

We can see this by working in components: if $u = (u_i), v = (v_i), x = (x_i), y = (y_i)$, then $\langle u \otimes x, v \otimes y \rangle = \sum_{i,j} u_i x_j v_i y_j$ which we may regroup into the product of the two inner products.

If u_1, \dots, u_n is an o.r. of G with handle c , and v_1, \dots, v_n is an o.r. of H with handle d , then $(u_i \otimes v_j)$ is an o.r. of $G \boxtimes H$, by the above identity.

Let $c \otimes d$ be a handle. Then

$$\begin{aligned} \theta(G \otimes H) &\leq \max_{u_i, v_j} \frac{1}{(\langle c \otimes d, u_i \otimes v_j \rangle)^2} \\ &= \max_{u_i, v_j} \frac{1}{(\langle c, u_i \rangle)^2 (\langle d, v_j \rangle)^2} \leq \theta(G) \theta(H). \end{aligned}$$

□

Lemma 10.14. $\theta(C_5) \leq \sqrt{5}$.

Proof. Start with

$$v_{i+1} = \frac{(\cos(\frac{2\pi}{5}i), (\sin(\frac{2\pi}{5}i), p))}{\sqrt{1+p^2}}$$

and increase p from zero until each $v_i \perp v_{i+2}, v_{i+3}$.

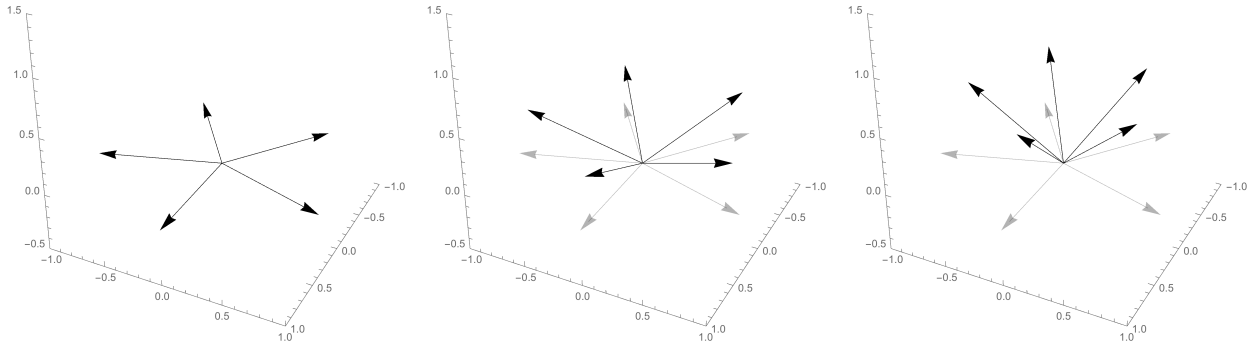


Figure 55: From left to right, the vectors v_i with $p = 0, 0.45, \frac{1+\sqrt{5}}{4} \approx 0.89$. The intuition is that we are starting with our five vectors equally spaced on the unit circle. Then we “close the umbrella”, increasing p , and pointing the vectors towards the z -axis, until each vector is orthogonal to the two vectors across from it.

We have $(1, 0, p) \perp (\cos \frac{4\pi}{5}, \sin \frac{4\pi}{5}, p)$ if $-p^2 = \cos \frac{4\pi}{5}$. Then $p^2 = \frac{1+\sqrt{5}}{4}$. Now, we may check that at this value of p , each $v_i \perp v_{i+2}, v_{i+3}$. That is, the v_i form an orthonormal representation of C_5 . Take $c = (0, 0, 1)$. Then $\langle c, v_i \rangle = \frac{p}{\sqrt{1+p^2}}$ for all i , and $\frac{1}{(\frac{p}{\sqrt{1+p^2}})^2} = \dots = \sqrt{5}$. □

References for Section 10.

- Alon, N. and Z. Füredi (1993). “Covering the Cube by Affine Hyperplanes”. In: *European Journal of Combinatorics* 14.2, pp. 79–83. ISSN: 0195-6698. DOI: <http://dx.doi.org/10.1006/eujc.1993.1011> (cit. on p. 79).
- Alon, N. and M. Tarsi (1989). “A nowhere-zero point in linear mappings”. In: *Combinatorica* 9.4, pp. 393–395. ISSN: 1439-6912. DOI: [10.1007/BF02125351](http://dx.doi.org/10.1007/BF02125351). URL: <http://dx.doi.org/10.1007/BF02125351> (cit. on p. 76).

- Da Silva, J. A. D. and Y. O. Hamidoune (1994). "Cyclic Spaces for Grassmann Derivatives and Additive Theory". In: *Bulletin of the London Mathematical Society* 26.2, pp. 140–146. DOI: [10.1112/blms/26.2.140](https://doi.org/10.1112/blms/26.2.140) (cit. on p. 78).
- Dvir, Zeev (2009). "On the size of Kakeya sets in finite fields". In: *Journal of the American Mathematical Society* 22.4, pp. 1093–1097. DOI: [10.1090/S0894-0347-08-00607-3](https://doi.org/10.1090/S0894-0347-08-00607-3) (cit. on p. 82).
- Erdős, P, A Ginzburg, and A Ziv (1961). "Theorem in the additive number theory". In: *Bulletin of the Research Council of Israel* 41-43 (cit. on p. 80).
- Hilbert, D. (1893). "Ueber die vollen Invariantensysteme". In: *Mathematische Annalen* 42.3, p. 320. ISSN: 1432-1807. DOI: [10.1007/BF01444162](https://doi.org/10.1007/BF01444162). URL: <http://dx.doi.org/10.1007/BF01444162> (cit. on p. 76).
- Kakeya, S. (1917). In: *Science Reports of the Tōhoku Imperial University*. First series, Mathematics, physics, chemistry 6, pp. 71–78 (cit. on p. 81).
- Lovász, L. (1979b). "On the Shannon capacity of a graph". In: *IEEE Transactions on Information Theory* 25.1, pp. 1–7. DOI: [10.1109/TIT.1979.1055985](https://doi.org/10.1109/TIT.1979.1055985) (cit. on p. 84).
- Erdős, P. and H. Heilbronn (1964). "On the addition of residue classes mod p ". eng. In: *Acta Arithmetica* 9.2, pp. 149–159 (cit. on p. 78).
- Shannon, C. (1956). "The zero error capacity of a noisy channel". In: *IRE Transactions on Information Theory* 2.3, pp. 8–19. DOI: [10.1109/TIT.1956.1056798](https://doi.org/10.1109/TIT.1956.1056798) (cit. on p. 82).

Collected references

- Ahlswede, R. and L. H. Khachatrian (Feb. 1997). "The Complete Intersection Theorem for Systems of Finite Sets". In: *Eur. J. Comb.* 18.2, pp. 125–136. ISSN: 0195-6698. DOI: [10.1006/eujc.1995.0092](https://doi.org/10.1006/eujc.1995.0092) (cit. on p. 20).
- Alon, N. and Z. Füredi (1993). "Covering the Cube by Affine Hyperplanes". In: *European Journal of Combinatorics* 14.2, pp. 79–83. ISSN: 0195-6698. DOI: <http://dx.doi.org/10.1006/eujc.1993.1011> (cit. on p. 79).
- Alon, N. and M. Tarsi (1989). "A nowhere-zero point in linear mappings". In: *Combinatorica* 9.4, pp. 393–395. ISSN: 1439-6912. DOI: [10.1007/BF02125351](https://doi.org/10.1007/BF02125351). URL: <http://dx.doi.org/10.1007/BF02125351> (cit. on p. 76).
- Bárány, I. (1982). "A generalization of Carathéodory's theorem". In: *Discrete Mathematics* 40.2, pp. 141–152 (cit. on p. 65).
- Bollobás, B. (1965). "On generalized graphs". In: *Acta Mathematica Academiae Scientiarum Hungarica* 16.3, pp. 447–452. ISSN: 1588-2632. DOI: [10.1007/BF01904851](https://doi.org/10.1007/BF01904851) (cit. on p. 9).
- Carathéodory, C. (1911). "Über den variabilitätsbereich der fourier'schen konstanten von positiven harmonischen funktionen". In: *Rendiconti del Circolo Matematico di Palermo (1884-1940)* 32.1, pp. 193–217. ISSN: 0009-725X. DOI: [10.1007/BF03014795](https://doi.org/10.1007/BF03014795) (cit. on p. 62).
- Conlon, D. (2009). "A new upper bound for diagonal Ramsey numbers". In: *Annals of Mathematics*, pp. 941–960 (cit. on p. 50).
- Da Silva, J. A. D. and Y. O. Hamidoune (1994). "Cyclic Spaces for Grassmann Derivatives and Additive Theory". In: *Bulletin of the London Mathematical Society* 26.2, pp. 140–146. DOI: [10.1112/blms/26.2.140](https://doi.org/10.1112/blms/26.2.140) (cit. on p. 78).
- de Caen, D. and Z. Füredi (2000). "The Maximum Size of 3-Uniform Hypergraphs Not Containing a Fano Plane". In: *Journal of Combinatorial Theory, Series B* 78.2, pp. 274–276. ISSN: 0095-8956. DOI: <http://dx.doi.org/10.1006/jctb.1999.1938> (cit. on p. 46).
- Dvir, Zeev (2009). "On the size of Kakeya sets in finite fields". In: *Journal of the American Mathematical Society* 22.4, pp. 1093–1097. DOI: [10.1090/S0894-0347-08-00607-3](https://doi.org/10.1090/S0894-0347-08-00607-3) (cit. on p. 82).
- Elekes, G. (1997). "On the number of sums and products". In: *Acta Arithmetica* 81.4, pp. 365–367 (cit. on pp. 4, 75).
- Erdős, P., A. Ginzburg, and A. Ziv (1961). "Theorem in the additive number theory". In: *Bulletin of the Research Council of Israel* 41-43 (cit. on p. 80).
- Furstenberg, H. and Y. Katznelson (1991). "A density version of the Hales-Jewett theorem". In: *Journal d'Analyse Mathématique* 57.1, pp. 64–119 (cit. on p. 56).
- Gowers, W. T. (2001). "A new proof of Szemerédi's theorem". In: *Geometric & Functional Analysis GAFA* 11.3, pp. 465–588. ISSN: 1420-8970. DOI: [10.1007/s00039-001-0332-9](https://doi.org/10.1007/s00039-001-0332-9) (cit. on p. 53).
- Green, B. and T. Tao (2008). "The primes contain arbitrarily long arithmetic progressions". In: *Annals of Mathematics* 167-2, pp. 481–547 (cit. on p. 57).
- Hales, A. W. and R. I. Jewett (1963). "Regularity and positional games". In: *Transactions of the American Mathematical Society* 106.2, pp. 222–229 (cit. on p. 54).
- Hall, P. (1935). "On Representatives of Subsets". In: *Journal of the London Mathematical Society* s1-10.1, pp. 26–30. DOI: [10.1112/jlms/s1-10.37.26](https://doi.org/10.1112/jlms/s1-10.37.26) (cit. on p. 6).
- Helly, E. (1923). "Über Mengen konvexer Körper mit gemeinschaftlichen Punkte." In: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 32, pp. 175–176. URL: <http://eudml.org/doc/145659> (cit. on p. 62).

- Hilbert, D. (1893). "Ueber die vollen Invariantensysteme". In: *Mathematische Annalen* 42.3, p. 320. ISSN: 1432-1807. DOI: [10.1007/BF01444162](https://dx.doi.org/10.1007/BF01444162). URL: <http://dx.doi.org/10.1007/BF01444162> (cit. on p. 76).
- Takeya, S. (1917). In: *Science Reports of the Tôhoku Imperial University*. First series, Mathematics, physics, chemistry 6, pp. 71–78 (cit. on p. 81).
- Katona, G. O. H. (1968). "Theory of Graphs: Proceedings of the Colloquium on Graph Theory, Held at Tihany, Hungary, September 1966". In: ed. by P. Erdős and G. O. H. Katona. Academic Press. Chap. A theorem of finite sets (cit. on p. 26).
- Katona, G. O. H., T. Nemetz, and M. Simonovits (1964). "On a graph problem of Turán." Hungarian. In: *Mat. Lapok*, pp. 228–238 (cit. on p. 31).
- Kleitman, D. J. (1970). "On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors". In: *Advances in Mathematics* 5.1, pp. 155–157. ISSN: 0001-8708. DOI: [10.1016/0001-8708\(70\)90038-1](https://dx.doi.org/10.1016/0001-8708(70)90038-1) (cit. on pp. 3, 14).
- Kruskal, J. B. (1963). "Mathematical optimization techniques". In: ed. by R. Bellman. University of California Press. Chap. The Number of Simplices in a Complex, p. 251 (cit. on p. 26).
- Liggett, T. M. (1977). "Extensions of the Erdős-Ko-Rado theorem and a statistical application". In: *Journal of Combinatorial Theory, Series A* 23.1, pp. 15–21. ISSN: 0097-3165. DOI: [10.1016/0097-3165\(77\)90075-9](https://dx.doi.org/10.1016/0097-3165(77)90075-9) (cit. on p. 21).
- Littlewood, J. E. and A. C. Offord (1938). "On the Number of Real Roots of a Random Algebraic Equation". In: *Journal of the London Mathematical Society* s1-13.4, pp. 288–295. DOI: [10.1112/jlms/s1-13.4.288](https://doi.org/10.1112/jlms/s1-13.4.288) (cit. on pp. 3, 12).
- Lovász, L. (1979a). *Combinatorial Problems and Exercises*. AMS/Chelsea publication. Problem 13.31. North-Holland Publishing Company. ISBN: 9780821869475 (cit. on p. 29).
- (1979b). "On the Shannon capacity of a graph". In: *IEEE Transactions on Information Theory* 25.1, pp. 1–7. DOI: [10.1109/TIT.1979.1055985](https://doi.org/10.1109/TIT.1979.1055985) (cit. on p. 84).
- Lubell, D. (1966). "A short proof of Sperner's lemma". In: *Journal of Combinatorial Theory* 1.2, pp. 299–. ISSN: 0021-9800. DOI: [10.1016/S0021-9800\(66\)80035-2](https://dx.doi.org/10.1016/S0021-9800(66)80035-2) (cit. on p. 9).
- Matoušek, J. (2002). *Lectures on discrete geometry*. Vol. 212. Graduate Texts in Mathematics. Springer. DOI: [10.1007/978-1-4613-0039-7](https://doi.org/10.1007/978-1-4613-0039-7) (cit. on p. 64).
- Meshalkin, L. D. (1963). "Generalization of Sperner's Theorem on the Number of Subsets of a Finite Set". In: *Theory of Probability & Its Applications* 8.2, pp. 203–204. DOI: [10.1137/1108023](https://doi.org/10.1137/1108023) (cit. on p. 9).
- Erdős, P. (Dec. 1945). "On a lemma of Littlewood and Offord". In: *Bull. Amer. Math. Soc.* 51.12, pp. 898–902. URL: <http://projecteuclid.org/euclid.bams/1183507531> (cit. on p. 12).
- (1947). "Some remarks on the theory of graphs". In: *Bulletin of the American Mathematical Society* 53.4, pp. 292–294 (cit. on p. 50).
- Erdős, P. and H. Heilbronn (1964). "On the addition of residue classes mod p ". eng. In: *Acta Arithmetica* 9.2, pp. 149–159 (cit. on p. 78).
- Erdős, P., C. Ko, and R. Rado (1961). "Intersection Theorems For Systems Of Finite Sets". In: *The Quarterly Journal of Mathematics* 12.1, pp. 313–320. DOI: [10.1093/qmath/12.1.313](https://doi.org/10.1093/qmath/12.1.313) (cit. on pp. 3, 17).
- Erdős, P. and M. Simonovits (1983). "Supersaturated graphs and hypergraphs". In: *Combinatorica* 3.2, pp. 181–192. ISSN: 1439-6912. DOI: [10.1007/BF02579292](https://doi.org/10.1007/BF02579292) (cit. on p. 44).
- Erdős, P. and A. H. Stone (Dec. 1946). "On the structure of linear graphs". In: *Bull. Amer. Math. Soc.* 52.12, pp. 1087–1091. DOI: [10.1090/S0002-9904-1946-08715-7](https://doi.org/10.1090/S0002-9904-1946-08715-7) (cit. on p. 39).
- Erdős, P. and G. Szekeres (1935). "A combinatorial problem in geometry". In: *Compositio Mathematica* 2, pp. 463–470 (cit. on pp. 49, 57, 70).

- Erdős, P. and E. Szemerédi (1983). "Studies in Pure Mathematics: To the Memory of Paul Turán". In: ed. by P. Erdős et al. Basel: Birkhäuser Basel. Chap. On sums and products of integers, pp. 213–218. ISBN: 978-3-0348-5438-2. DOI: [10.1007/978-3-0348-5438-2_19](https://doi.org/10.1007/978-3-0348-5438-2_19) (cit. on pp. 4, 74).
- Rado, R. (1943). "Note on combinatorial analysis". In: *Proceedings of the London Mathematical Society* 48.2, pp. 122–160 (cit. on p. 54).
- Radon, J. (1921). "Mengen konvexer Körper, die einen gemeinsamen Punkt enthalten". In: *Mathematische Annalen* 83.1, pp. 113–115. ISSN: 1432-1807. DOI: [10.1007/BF01464231](https://doi.org/10.1007/BF01464231) (cit. on p. 61).
- Ramsey, F.P. (1930). "On a Problem of Formal Logic". In: *Proceedings of the London Mathematical Society* 2.1, pp. 264–286 (cit. on p. 49).
- Sarkaria, K. S. (1992). "Tverberg's theorem via number fields". In: *Israel journal of mathematics* 79.2, pp. 317–320 (cit. on p. 66).
- Shannon, C. (1956). "The zero error capacity of a noisy channel". In: *IRE Transactions on Information Theory* 2.3, pp. 8–19. DOI: [10.1109/TIT.1956.1056798](https://doi.org/10.1109/TIT.1956.1056798) (cit. on p. 82).
- Soifer, A., ed. (2010). *Ramsey Theory: Yesterday, Today, and Tomorrow*. Vol. 285. Progress in Mathematics. Birkhäuser Boston. ISBN: 0817680918. DOI: [10.1007/978-0-8176-8092-3](https://doi.org/10.1007/978-0-8176-8092-3) (cit. on p. 54).
- Solymosi, J. (2009). "Bounding multiplicative energy by the sumset". In: *Advances in Mathematics* 222.2, pp. 402–408. ISSN: 0001-8708. DOI: <http://dx.doi.org/10.1016/j.aim.2009.04.006> (cit. on p. 4).
- Spencer, J. (1975). "Ramsey's theorem—A new lower bound". In: *Journal of Combinatorial Theory, Series A* 18.1, pp. 108–115. ISSN: 0097-3165. DOI: [http://dx.doi.org/10.1016/0097-3165\(75\)90071-0](http://dx.doi.org/10.1016/0097-3165(75)90071-0) (cit. on p. 50).
- Sperner, E. (1928). "Ein Satz über Untermengen einer endlichen Menge". In: *Mathematische Zeitschrift* 27.1, pp. 544–548. ISSN: 1432-1823. DOI: [10.1007/BF01171114](https://doi.org/10.1007/BF01171114) (cit. on p. 8).
- Steiner, J. (1838). "Einfache Beweise der isoperimetrischen Hauptsätze." In: *Journal für die reine und angewandte Mathematik* 18, pp. 281–296 (cit. on p. 23).
- Szemerédi, E. (1975). "On sets of integers containing k elements in arithmetic progression". In: *Acta Arithmetica* 27.1, pp. 199–245 (cit. on p. 57).
- Szemerédi, E. and W. T. Trotter (1983). "Extremal problems in discrete geometry". In: *Combinatorica* 3.3, pp. 381–392. ISSN: 1439-6912. DOI: [10.1007/BF02579194](https://doi.org/10.1007/BF02579194) (cit. on pp. 4, 74).
- Turán, P. (1941). "On an extremal problem in graph theory". In: *Mat. Fiz. Lapok* 48.436-452, p. 137 (cit. on p. 33).
- Tverberg, H. (1966). "A Generalization of Radon's Theorem". In: *Journal of the London Mathematical Society* s1-41.1, pp. 123–128. DOI: [10.1112/jlms/s1-41.1.123](https://doi.org/10.1112/jlms/s1-41.1.123) (cit. on p. 66).
- van der Waerden, B. L. (1927). "Beweis einer Baudetschen Vermutung." German. In: *Nieuw Arch. Wiskd., II. Ser.* 15, pp. 212–216. ISSN: 0028-9825 (cit. on p. 50).
- Wit, Ernst (1952). "Ein kombinatorischer Satz der Elementargeometrie". In: *Mathematische Nachrichten* 6.5, pp. 261–262. DOI: [10.1002/mana.19520060502](https://doi.org/10.1002/mana.19520060502) (cit. on p. 54).
- Yamamoto, K. (Dec. 1954). "Logarithmic order of free distributive lattice". In: *J. Math. Soc. Japan* 6.3-4, pp. 343–353. DOI: [10.2969/jmsj/00630343](https://doi.org/10.2969/jmsj/00630343) (cit. on p. 9).

Index

- k -coloring, 38
- affine
 - dependence, 60
 - hull, 60
 - subspace, 60
- arithmetic progression, 50
- centerpoint, 63
- chain, 8
- colexicographic order, 26
- colorful simplex, 65
- combinatorial line, 53
- Combinatorial Nullstellensatz, 76
- compressed, 23
- compression, 23
- convex, 61
 - combination, 61
 - dependence, 61
 - hull, 61
 - independence, 61
- density of graph, 34
- graph
 - bipartite, 6
 - matching, 6
 - neighborhood, 6
- hypergraph, 5
- hypergraph Lagrangian, 42
- incidences, 4
- initial segment of A , 27
- intersecting set system, 2, 17
- isomorphic, 31
- Keakeya set, 82
- Lagrangian, 34
- lexicographic order, 26
- linear
 - dependence, 60
 - hull, 60
 - space, 60
- link graph, 46
- Littlewood-Offord
 - sets of numbers, 12
 - sets of vectors, 14
- Lovász theta, 84
- multigraph, 46
- orthonormal representation, 84
 - value, 84
- polychromatic m -tuple, 51
- power set, 5
- Ramsey number, 49
- regular subsystem, 49
- root, 53
- set system, 5
- shadow, 10
- sparse
 - set of indices of integers modulo p , 15
 - set of indices of vectors, 14
- Sperner system, 3
- strong k -coloring, 42
- subsequence
 - decreasing, 57
 - increasing, 57
- symmetric
 - chain, 12
 - partition, 14
- system of distinct representatives, 7
- t -intersecting, 18
- Turán density, 31
- Turán graph, 33
- Turán number, 31
- Van der Waerden number, 50