Lecture 12. Extremal set theory, part 4
Restricted intersection and dimension argument

Hong Liu

25th May 2021

Today, we will consider set systems with restricted intersections and introduce a dimension argument, which is very useful when the extremal problem has many distinct extremal structures. The idea is that, when counting the number of objects in a discrete family $\mathcal{F}$, if we can map $\mathcal{F}$ injectively into a vector space such that the image is linearly independent, then the dimension of the space bounds the size of $\mathcal{F}$ from above.

For a set A, we write $\mathbb{1}_A$ for the indicator function/characteristic vector of A.

1 Odd town even town

Let us start with a basic but classical example of dimension argument of odd town even town problem.

In a town of n residents, people want to form clubs such that any two different clubs have even number of common members. If each club must be of also even size, then how many clubs there can be? A moment of thoughts yields that we can have exponentially many: pair up elements $\{1,2\},\{3,4\},\ldots,\{2\lfloor n/2\rfloor-1,2\lfloor n/2\rfloor\}$ and then we can let clubs C_i be all possible combinations of pairs, for which there are $2^{\lfloor n/2\rfloor}$ many choices.

What if we require instead that each club has odd size? We can then take $C_i=\{i\}$, $i\in[n]$, to be the singletons, or if n is even, take $C_i=[n]\setminus\{i\}$. Though these examples are only of linear size. Somewhat surprisingly, this is the best we can do under such restrictions.

Theorem 1.1. *Let $\mathcal{F}\subseteq 2^{[n]}$ be a family of sets such that*

- *for each $F\in\mathcal{F}$, $|F|=1 \mod 2$, and*
- *for distinct $F,F'\in\mathcal{F}$, $|F\cap F'|=0 \mod 2$.*

Then $|\mathcal{F}|\le n$.

Proof. Associate $F\in\mathcal{F}$ with $\mathbb{1}_F\in\mathbb{F}_2^n$. Note that, as for any $F,F'\in\mathcal{F}$, $\langle\mathbb{1}_F,\mathbb{1}_{F'}\rangle=1$ (in $\mathbb{F}_2$) if and only if $F=F'$, $\mathbb{1}_F$, $F\in\mathcal{F}$, form orthonormal basis of $\mathbb{F}_2^n$, which has dimension n. Thus $|\mathcal{F}|\le n$ as desired. □

2 Sets with at most two distinct distances

Sometimes additional tricks are available to further reduce the dimension of the space when we do dimension argument. Let us consider the following problem.

Given a set $X\subseteq\mathbb{R}^n$ of points, if there is only one distinct distance between pairs of points in X, then X is a regular simplex, which has size at most $n+1$ in $\mathbb{R}^n$. What if we are allowed to have two distinct distances? First, for a lower bound, we can take X to be e_i+e_j, $ij\in\binom{[n]}{2}$, i.e. sums of pairs of standard basis vectors. This set has size $\binom{n}{2}$. It is an open problem to find the exact maximum size of such set. The best current upper bound, due to Blokhuis, is $\binom{n+2}{2}$. We prove here a slightly weaker bound.

Theorem 2.1. *Let $X \subseteq \mathbb{R}^n$ be a set with at most two distinct distances between pairs of its points, then $|X| \le \frac{1}{2}(n+4)(n+1)$.*

Proof. Let $X = \{v_1, \dots, v_t\}$ and let d, D be the two distinct distances. Associate to each point $v_i \in X$ the n-variate degree-4 polynomial P_i defined as follows: for $x \in \mathbb{R}^n$,

$$P_i(x) = \left(\|x - v_i\|^2 - d^2\right)\left(\|x - v_i\|^2 - D^2\right)^1.$$

We claim that P_i, $i \in [t]$, are linearly independent. It follows from that $P_i(v_j)$ is 0 if $i = j$ and d^2D^2 otherwise. Indeed, if $\sum_i \alpha_i P_i = 0$, then for any $j \in [t]$, we have

$$0 = \sum_i \alpha_i P_i(v_j) = \alpha_j P_j(v_j) = d^2 D^2 \alpha_j,$$

implying that $\alpha_j = 0$. This yields a $O(n^4)$ upper bound on $|X|$. We can however do much better by choosing a more economic set of vectors spanning P_is.

Note that $\|x - v_i\|^2 = \sum_{j\in[n]} x_j^2 - 2\sum_{j\in[n]} x_j v_{ij} + \sum_{j\in[n]} v_{ij}^2$. Thus, each P_i is a linear combination of the following polynomials:

- $\left(\sum_{j\in[n]} x_j^2\right)^2$;
- $x_k \cdot \sum_{j\in[n]} x_j^2$, $k \in [n]$ (n of them);
- $x_j \cdot x_k$, $1 \le j \le k \le n$ ($(n+1)n/2$ of them);
- x_j, $j \in [n]$ (n of them);
- the constant term 1.

Thus the dimension of the space containg all P_i is at most $1 + n + (n+1)n/2 + n + 1 = (n+4)(n+1)/2$, which upper bounds $t = |X|$ as P_is are linearly independent. $\square$

3 Families with restricted intersections

There are 2^n subsets of $[n]$. We have seen that if we forbid empty intersection (between pairs of sets in a family), then the family can have size at most $\frac{1}{2} \cdot 2^n$, with equality holds for families whose sets contain a common elements. Even in the uniform case, we could still have k-uniform intersecting families of size $\binom{n-1}{k-1}$; when k is close to $n/2$, such families are still quite large. Is there a restriction of a single forbidden size between pair intersections that would drastically change how large a family can be, in particular, forcing the family to have size expenentially smaller than 2^n? As we shall see soon that, the answer is yes.

Theorem 3.1 (Frankl-Wilson 81)**.** *Let p be an odd prime and $\mathcal{F} \subseteq 2^{[n]}$. If*

- *for any $F \in \mathcal{F}$, $|F| = 0 \mod p$;*
- *for any distinct $F, F' \in \mathcal{F}$, $|F \cap F'| \ne 0 \mod p$,*

then

$$|\mathcal{F}| \le \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1}.$$

[1] Here we use ℓ_2-norm, i.e. $\|u\| = \sqrt{\sum_i u_i^2}$.

Proof. We associate to each set $F \in \mathcal{F}$ an n-variate degree-$(p-1)$ polynomial P_F over the field $\mathbb{F}_p$ as follows: for $x \in \mathbb{F}_p^n$,

$$P_F(x) = 1 - \Big(\sum_{i\in F} x_i\Big)^{p-1}.$$

Observe first that the polynomials P_F, $F \in \mathcal{F}$, are linearly independent. To see this, take $F, F' \in \mathcal{F}$, then, as $|F \cap F'| = 0 \mod p$ if and only if $F = F'$, we have by Fermat's little theorem that

$$P_F(\mathbb{1}_{F'}) = 1 - |F \cap F'|^{p-1} = \begin{cases} 1, & \text{if } F = F'; \\ 0, & \text{otherwise.} \end{cases}$$

We need one more trick to further reduce the dimension of the space.

Multilinearisation. Note that when we prove that P_F, $F \in \mathcal{F}$, are linearly independent, we evaluate them over $\{0,1\}$-vectors. Since over $\{0,1\}$-vectors, any monomial say $x_1^3x_2^5x_3^2$ takes the same value as $x_1x_2x_3$. We can then *multilinearise* P_F by replacing all powers x_i^a, $a \geq 2$ in P_F, by x_i.

Finally, the dimension of the space of multilinear n-variate degree-$(p-1)$ polynomials is at most the number of such monomials, which is

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1},$$

as desired. □

The following special case of the above theorem is quite surprising. Consider $n = 4p$, then a $2p$-uniform set system can have size up to $\binom{n}{2p} \sim \frac{1}{\sqrt{2\pi p}} \cdot 2^n$. But if we forbid intersections of size p, then Theorem 3.1 implies that the family has to be exponentially small: at most $\sum_{i=0}^{p-1}\binom{n}{i}$.

4 Sets with no orthogonal pairs are exponentially small

A nice geometric corollary of Theorem 3.1 is the following, implying that forbidding orthogonal pairs in a set of vectors forces the set to be exponentially small. The idea is to discretise the problem using (random) averaging; in the discrete set (L below) of unit vectors that we shall use, orthogonal pairs correspond to pairs with forbidden intersection size.

Corollary 4.1. *Let p be an odd prime, $n = 4p$ and $V \subseteq \mathbb{S}^{n-1}$ be a set of unit vectors in $\mathbb{R}^n$. If no pairs in V are orthogonal, then the measure[2] $\mu(V)$ is exponentially small.*

Proof. Let $\mathcal{F} = \binom{[n]}{n/2} = \binom{[4p]}{2p}$ be the family of all $2p$-subsets of $[n]$, and $\alpha = \frac{1}{\sqrt{n}}$. Associate to each $F \in \mathcal{F}$ a unit vector v_F that takes value α on coordinates in F and $-\alpha$ otherwise, i.e.

$$v_F = \alpha\big(\mathbb{1}_F - \mathbb{1}_{[n]\setminus F}\big).$$

Set

$$L = \{v_F : F \in \mathcal{F}\}.$$

Note that for any $F, G \in \mathcal{F}$,

$$\langle v_F, v_G\rangle = \alpha^2\big(|F\cap G| + |[n]\setminus(F\cup G)| - |F\triangle G|\big) = \frac{1}{n}\big(n - 2|F\triangle G|\big) = 1 - \frac{1}{2p}|F\triangle G|. \quad (1)$$

Thus, $\langle v_F, v_G\rangle = 0$ if and only if $|F\cap G| = p$. Thus by Theorem 3.1, $|V \cap L| \leq \sum_{i=0}^{p-1}\binom{n}{i}$, which is exponentially smaller[3] than 2^n.

[2] The normalised spherical measure, which is rotation-invariant.
[3] Using e.g. stirling's formula

We now do an averaging. Take a uniform random rotation ρ. Then any unit vector has probability $\mu(X)$ to fall in ρV. Thus,

$$\mathsf{E}|\rho V \cap L| = \mu(X) \cdot |L| = \mu(X) \cdot \binom{n}{n/2}.$$

On the other hand, for any rotation ρ, ρV has no orthogonal pairs either, and so as above we have $|\rho V \cap L| \leq \sum_{i=0}^{p-1} \binom{n}{i}$. Solving $\mathsf{E}|\rho V \cap L| = \mu(X) \cdot \binom{n}{n/2} \leq \sum_{i=0}^{p-1} \binom{n}{i}$ yields the desired bound. $\square$

5 Disproof of Borsuk's conjecture by Kahn and Kalai

Borsuk in 1933 conjectured that every bounded set in $\mathbb{R}^n$ can be decomposed into at most $n+1$ sets of smaller diameter. This old conjecture was disproved by Kahn and Kalai in 1993 in a strong sense: if we want to cut a set into smaller pieces, then exponentially many pieces are needed!

Their proof is an elegant use of Theorem 3.1 as follows. By taking tensor, we can construct a set in which the orthogonal pairs are the ones that are furthest apart, and so any subset of smaller diameter has to avoid orthogonal pairs which then has to be exponentially small.

Corollary 5.1. *Let p be an odd prime and $n = 4p$. Then there exists a set $X \subseteq \mathbb{R}^{n^2}$ of size $\binom{n}{n/2}$ such that every subset of X with smaller diameter has size at most $\sum_{i=0}^{p-1} \binom{n}{i}$.*

Proof. As in Corollary 4.1, define

$$\mathcal{F} = \binom{[n]}{n/2} \quad \text{and} \quad L = \Big\{ v_F = \frac{1}{\sqrt{n}} \big(\mathbb{1}_F - \mathbb{1}_{[n]\setminus F} \big) : F \in \mathcal{F} \Big\}.$$

Then let $X \subseteq \mathbb{R}^{n^2}$ be the set of rank-1 outer products of vectors[4] in L:

$$X = \{ v \otimes v : v \in L \}.$$

Note that

$$\langle v \otimes v, w \otimes w \rangle = \sum_{i,j} v_i v_j w_i w_j = \sum_i v_i w_i \cdot \sum_j v_j w_j = \langle v, w \rangle^2 \geq 0$$

with equality if and only if v, w are orthogonal, and that $v \otimes v \in \mathbb{R}^{n^2}$ is a unit vector when $v \in \mathbb{R}^n$ is a unit vector. Consequently,

$$\|v \otimes v - w \otimes w\|^2 = \|v \otimes v\|^2 + \|w \otimes w\|^2 - 2\langle v \otimes v, w \otimes w \rangle = 2 - 2\langle v, w \rangle^2 \leq 2$$

with equality if and only if v, w are orthogonal. Thus, the diameter of X is $\sqrt{2}$, and any subset of X with smaller diameter contains no pairs $v \otimes v, w \otimes w$ where v, w are orthogonal.

Finally, we have seen in (1) that v_F, v_G are orthogonal if and only if $|F \cap G| = p$. Thus by Theorem 3.1, any subset of X of diameter smaller than $\sqrt{2}$ has size at most $\sum_{i=0}^{p-1} \binom{n}{i}$. $\square$

It follows right away that to cover the above set X using pieces of smaller diameter, we need at least $\binom{n}{n/2} / \sum_{i=0}^{p-1} \binom{n}{i} = e^{\Omega(n)}$ many pieces.

[4] We view the n-by-n matrix $v \otimes v = v \cdot v^T$ here as a length-n^2 vector in $\mathbb{R}^{n^2}$.