18.218: Algebraic Methods in Extremal Combinatorics

Lecturer: Professor Lisa Sauermann

Notes by: Andrew Lin

Spring 2022

Introduction

The topic for this semester is **algebraic methods in extremal combinatorics**. Lectures for this class will be recorded in case any of us need to miss class (for COVID or other reasons), but we should still make sure to attend in person. If we need to miss class due to COVID isolation, we should contact the course staff or find classmates who are taking notes. (And if we're feeling sick or need help for any reason, we should contact S^3 or GradSupport, who will advocate on our behalf.)

Logistically, office hours will be held in 2-171 from 1-2 on Mondays, and grading will be done using five (biweekly) problem sets. Homework will always appear on Canvas, and submission should be done on Gradescope. (But if this causes any problems for us, we should email Professor Sauermann and we can figure something out). Collaboration is encouraged, but we should think about the problems ourselves, and we should only write down a solution if it came out of a discussion that we were actively involved in (in other words, we can't ask "what is the solution"). We should always make sure to indicate our collaborators at the start of a solution.

The syllabus also includes a schedule and list of topics which is subject to change, but we can take a look on Canvas to see the specifics there. There are no required references, so in theory we should be fine just showing up to lecture and learning here!

Fact 1

Professor Sauermann had the students introduce ourselves by name, year, and major.

1 January 31, 2022

In the first three weeks of this class, we'll discuss **methods in linear algebra**, starting with a classic example:

Problem 2

Consider a town of n citizens who form (potentially intersecting) clubs, subject to the following rules:

- Each club has an odd number of members.
- Any two clubs have an even number of members in common.

What is the maximum number of clubs we can form in this town (as a function of n)?

One potential situation could be that we have n clubs, and each citizen is in a different club (this satisfies the conditions above). But if we try to start forming larger groups, we start running into problems – even if we try to form clubs of size 3 to get more than n total clubs, it's hard to come up with constructions without having two clubs that intersect with just a single person. It turns out that we can't actually improve the simple construction that gets us n clubs:

Theorem 3 (Odd-town theorem)

Under the assumptions of Problem 2, there cannot be more than n clubs. Removing the flavor text, if we have (distinct) subsets $C_1, \cdots, C_m \subseteq \{1, \cdots, n\}$ such that $|C_i|$ is odd and $|C_i \cap C_j|$ is even for all $1 \le i \ne j \le n$, then $m \le n$.

Proof. (We have already achieved $m = n$ by construction.) We'll first transform the setup into a linear algebraic one. Number the citizens of the town from 1 to n, and for each club, form a vector in $\{0, 1\}^n$, such that the cth entry of the vector is 1 if citizen c is a member of the club (this is known as the **incidence vector** of the club). If we have m clubs, then we get m vectors $v_1, \cdots, v_m$ in $\{0, 1\}^n$, and we wish to show that $m \le n$.

Since each club has an odd number of members, each v_i has an odd number of ones, and since any two clubs have an even number of common members, $\mathbf{v_i} \cdot \mathbf{v_j}$ **must be even** for all i, j (because this dot product counts the number of indices where v_i and v_j are both 1). Motivated by that, notice that we can restate our first condition as $\mathbf{v_i} \cdot \mathbf{v_i}$ **being odd** for all i.

Since dealing with even and odd-ness can be annoying, we can instead interpret $v_1, \cdots, v_m$ as vectors in the finite field $\mathbb{F}_2^n$, so that $v_i \cdot v_j$ is 1 if $i = j$ and 0 otherwise. And thus it suffices to show that the v_is are **linearly independent**, from which the result must follow (by finite-dimensional linear algebra). Indeed, suppose $\lambda_1 v_1 + \cdots + \lambda_m v_m = 0$ for some coefficients $\lambda_1, \cdots, \lambda_m \in \mathbb{F}_2$. Then taking the dot product of both sides with v_i, most terms disappear because of the orthogonality we described above. So we find $\lambda_i v_i \cdot v_i = 0$, so $\lambda_i = 0$ for all i, which is what we need to show linear independence, and thus we can only have $\dim(\mathbb{F}_2^n) = n$ total clubs. □

Extermal combinatorics is a field which studies questions of the form "given some configuration of graphs or sets or other combinatorial objects, under certain conditions, what is the maximum or minimum size of a configuration?" So the result we've just proved is a prototypical example of the kind of results we'll be seeing in this class.

Problem 4

We'll now ask what happens if we adjust the conditions of our problem statement a little bit, replacing "odd" in the odd-town theorem by "even." We still wish to find out the maximum number of clubs that this town may have.

The construction from above now no longer works, but we can still have a variation of this construction: if we take disjoint sets of size 2, we can get $\frac{n}{2}$ sets (and in extermal combinatorics we often don't really care about constant factors like 2, since asymptotic considerations are often the best we can do). But this time we can do better – if we take pairs of pairs, we still satisfy the conditions in the problems, and in fact we can take any set of these pairs and that will still work. (In other words, imagine that we have $\lfloor \frac{n}{2} \rfloor$ married couples, and each club either contains the couple or neither member.) Then all $2^{\lfloor n/2 \rfloor}$ of these potential clubs will be of even size and have even intersection (because we always just have a set of couples in each case), and we can notice that this small change to the odd-town problem statement has led us to a dramatically different answer!

Theorem 5 (Even-town theorem)

Suppose we have distinct subsets $C_1, \cdots, C_m \subseteq \{1, \cdots, n\}$ such that $|C_i|$ and $|C_i \cap C_j|$ are even for all $1 \le i \neq j \le n$. Then $m \le 2^{\lfloor n/2 \rfloor}$.

Proof. (We have already achieved $m = 2^{\lfloor n/2 \rfloor}$ by construction.) We again form the incidence vectors $v_1, \cdots, v_m \in \mathbb{F}_2^n$ of the sets $C_1, \cdots, C_m$, in the same way as before. We may consider the **span** of these vectors $U = \text{span}(v_1, \cdots, v_m)$; we can verify that we can always add new subsets in this span U and still preserve the theorem conditions, but we don't need that result directly. Instead, we can just observe that $m \le |U| = 2^{\dim U}$ (because m is the number of vectors in $(v_1, \cdots, v_m)$, while $|U|$ is the number of vectors in its span), so it suffices to show that $\dim U \le \lfloor n/2 \rfloor$.

This time, we have $v_i \cdot v_j = 0$ for all i, j by the theorem statement. Notice that for any $u, u' \in U$, we have $u \cdot u' = 0$, because we can write out $u = \lambda_1 v_1 + \cdots + \lambda_m v_m$ and $u' = \lambda_1' v_1 + \cdots + \lambda_m' v_m$ and find that (by distributivity)

$$u \cdot u' = \sum_{i,j=1}^{n} \lambda_i' \lambda_j v_i \cdot v_j = \sum_{i,j=1}^{n} \lambda_i' \lambda_j \cdot 0 = 0,$$

where this calculation is being done in $\mathbb{F}_2$. (This means that U is known as a **totally isotropic subspace**.) We can now consider the dual space $\mathcal{L}(\mathbb{F}_2^n)$, which is the set of all **linear forms** on $\mathbb{F}_2^n$ (that is, the set of linear maps $\mathbb{F}_2^n \to \mathbb{F}_2$). This vector space is also of dimension n, just like $\mathbb{F}_2^n$. Now for any vector $v \in \mathbb{F}_2^n$, consider its corresponding linear form $\phi_v \in \mathcal{L}(\mathbb{F}_2^n)$ defined by

$$\phi_v(x) = v \cdot x.$$

(In other words, the linear form ϕ_v is the "dotting with v" map.) Sending v to ϕ_v can then be represented as a map $\Phi : \mathbb{F}_2^n \to \mathcal{L}(\mathbb{F}_2^n)$, and in fact (we can check mechanically that) this Φ is itself a linear map.

In fact, Φ is an isomorphism – to show this, because we already know that $\mathbb{F}_2^n$ and $\mathcal{L}(\mathbb{F}_2^n)$ are vector spaces of dimension n, we just need to show injectivity. Indeed, if Φ mapped some vector v to the zero map, that would mean that $\phi_v(x) = v \cdot x = 0$ for all $x \in \mathbb{F}_2^n$. But whenever v has some nonzero coordinate i, we get a contradiction by picking x to be the ith standard basis vector. Thus Φ has a trivial kernel and is injective.

Remark 6. *We may have been told in our linear algebra classes that there is no canonical isomorphism between a vector space and its dual (only a vector space and its double dual). While mapping $v \to \phi_v$ may look very natural, this map Φ is* **not canonical** *because it depends on an inner product, which depends on some choice of basis.*

We're now ready to return to the proof: let W be the subspace of $\mathcal{L}(\mathbb{F}_2^n)$ consisting of all linear forms $\phi \in \mathcal{L}(\mathbb{F}_2^n)$ such that $\phi(u) = 0$ for all $u \in U$. We claim that $\Phi(U) \subseteq W$; indeed, for any $u' \in U$, we wish to show that $\phi_{u'} \in W$, and that's true because $\phi_{u'}(u) = u' \cdot u = 0$ by our earlier calculation. But because Φ is an isomorphism that maps U into W, we know that $\boxed{\dim U \le \dim W}$, but furthermore W is the kernel of the linear map $\mathcal{L}(\mathbb{F}_2^n) \to \mathcal{L}(U)$ defined by restriction (in other words, taking any linear form ϕ and only looking at its action $\phi|_U$ on U). Furthermore, this linear map is surjective, because we can extend any linear form on U to one on $\mathbb{F}_2^n$ (pick a basis of U and then complete that basis), so the rank-nullity theorem tells us that

$$\dim W = \dim(\mathbb{F}_2^n) - \dim \mathcal{L}(U) = n - \dim U.$$

Thus plugging this back into the boxed inequality above, using the fact that $\dim U$ is an integer,

$$\dim U \le n - \dim U \implies 2 \dim U \le n \implies \dim U \le \lfloor \frac{n}{2} \rfloor,$$

which was the result that we wished to show. □

2 February 3, 2022

Last time, we discussed the odd-town and even-town theorem, which described the maximum number of (odd and even, respectively)-size subsets on $\{1, 2, \cdots, n\}$, where any two subsets have an even-size intersection. We found that this maximum was n and $2^{\lfloor n/2 \rfloor}$, respectively, but in both cases we used linear algebra to show the result.

The next natural extension is to ask any two subsets to have an odd-size intersection:

Problem 7

Consider distinct subsets $C_1, \cdots, C_m \subseteq \{1, 2, \cdots, n\}$, such that $|C_i|$ is even for all $1 \le i \le m$ and $|C_i \cap C_j|$ is odd for all $i \ne j$. What is the maximum possible value of m?

A simple construction we can consider is $C_1 = \{1, 2\}, C_2 = \{1, 3\}, C_3 = \{1, 4\}, \cdots$, which gives us $m = n - 1$. (In each case, the intersection $C_i \cap C_j$ only contains 1.) Furthermore, we can also add the set $C_n = \{2, 3, \cdots, n\}$ as long as n is odd. But that's actually the best we can do:

Theorem 8

Under the settings of Problem 7, the maximum value of m is n for odd n and $n - 1$ for even n.

This separation between odd and even n may seem surprising to us (linear algebra "shouldn't be able to tell" the parity of n), but we'll see in a moment where this comes up in the proof.

Proof. (We have achieved both equality cases by construction.) **First, we'll do the even n case**, and we'll again use the same linear algebra setup that we've been doing with our earlier proofs. Let $v_1, \cdots, v_m \in \mathbb{F}_2^n$ be the incidence vectors of the sets $C_1, \cdots, C_m$, respectively; the theorem statement requires that $v_i \cdot v_j = 0$ if $i = j$ and $v_i \cdot v_j = 1$ otherwise (remember that dot products count the number of common 1s in the two vectors).

Notice that this time we need to prove $m \le n - 1$, so showing linear independence isn't enough on its own. But in this case we only need a little more, essentially because **the vectors also can't span the whole space**. In particular, if we consider the subspace $U = \{(x_1, \cdots, x_n) \in \mathbb{F}_2^n : x_1 + \cdots + x_n = 0\}$, which is clearly a linear subspace, notice that U contains the set of vectors with an even number of 1s, and thus this is also the set of vectors u such that $u \cdot u = 0$. So because $v_1, \cdots, v_m$ are all in U, their span is contained in U, which is strictly smaller than $\mathbb{F}_2^n$. Thus the span is of dimension at most $\dim U \le n - 1$, and now it suffices to show linear independence.

Indeed, suppose for the sake of contradiction that $m \ge n$. Because $\dim U = n - 1$, these vectors must be linearly dependent – consider the first n of these vectors. Then we can write

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$$

for some $\lambda_1, \cdots, \lambda_n \in \mathbb{F}_2$ not all zero. If we now take the dot product with v_j, we find that

$$\sum_i \lambda_i v_i \cdot v_j = \sum_{i \ne j} \lambda_i = 0,$$

or equivalently adding λ_j back in, we have

$$\lambda_1 + \cdots + \lambda_n = \lambda_j.$$

Since this holds for all λ_j, they must all be equal, and because we assumed they're all not zero, we must have $\lambda_j = 1$ for all j. But this is a contradiction if n is even (one side is even and the other is odd), so indeed we can only have $m \le n - 1$ in this case, as desired.

The case where n is odd can now follow in a variety of ways: we can either replace n with $n+1$ in the linear independence argument, or we can replace each incidence vector with its complement and apply the odd-town theorem from last lecture. But a third method is to take the clubs $C_i \subseteq \{1, 2, \cdots, n\}$ and consider them as subsets of $\{1, 2, \cdots, n+1\}$ instead. That doesn't change the size of sets and their intersections (we're just adding a citizen who doesn't participate in any clubs), and then by the even case we have $m \le (n+1) - 1 = n$. □

The last case, where both $|C_i|$ and $|C_i \cap C_j|$ are odd, is left as an exercise to our homework (which will be posted next Tuesday). We'll just say one more thing about the odd-town theorem before we move on: if we return to the classical odd-town theorem, where $|C_i|$ is odd and $|C_i \cap C_j|$ is even, we can actually restate the proof in a slightly different way:

Alternative proof of the odd-town theorem. As before, let $v_1, \cdots, v_m \in \mathbb{F}_2^n$ be the incidence vectors of $C_1, \cdots, C_m$, where $v_i \cdot v_j = 1$ if $I = j$ and 0 otherwise. The dot product $v_i \cdot v_j$ can also be written as the matrix multiplication $v_i^T v_j$ (where we treat each v_j as an $n \times 1$ matrix). So if we define the $n \times m$ matrix A with columns $v_1, \cdots, v_m$, then $A^T A$ has (i, j)th entry $v_i^T v_j$, so $\boxed{A^T A = I_m}$ (where I_m denotes the identity matrix). But this means that $m = \text{rank}(I_m) = \text{rank}(A^T A) \le \text{rank}(A)$ (since the rank of a product is at most the rank of each term), and that rank is at most n because A is an $n \times m$ matrix. Thus $m \le n$ as desired. □

Presenting this proof is also meant to motivate the next result that we're presenting:

Theorem 9 (Fisher's inequality)

Let $C_1, \cdots, C_m \subseteq \{1, 2, \cdots, n\}$ be distinct nonempty subsets, such that all pairwise intersections $C_i \cap C_j$ (for $i \ne j$) have the same size. Then $m \le n$.

Proof. We can achieve $m = n$ either by setting $C_i = \{i\}$, or using the construction from Problem 7, or using all subsets of size $(n-1)$. To show the inequality, first say that $|C_i \cap C_j| = t$ for some integer t, meaning that $|C_i| \ge t$ for all i. We split into two cases (assuming $m \ge 2$, since $m = 0, 1$ is clear):

- Suppose $|C_j| = t$ for some j – without loss of generality, suppose $j = 1$, and notice that in this case $t \ge 1$ because our subsets are nonempty. Then each of the other sets C_i must contain C_1 so that $C_i \cap C_1$ contains t elements, and in fact if $C_1 \subseteq C_\ell, C_h$, then
$$C_1 \subseteq C_\ell \cap C_h \quad \text{for all } 2 \le \ell < h \le m.$$
Now because C_1 and $C_\ell \cap C_h$ both have t elements, that means $C_\ell \cap C_h = C_1$ for all $2 \le \ell < h \le m$. Thus the $(m-1)$ sets $C_2 \cap C_1, \cdots, C_m \cap C_1$ are **disjoint** subsets of the set $\{1, 2, \cdots, n\} \setminus C_1$, and they are nonempty because the C_is are distinct. Thus there are at most $|\{1, 2, \cdots, n\} \setminus C_1| = n - |C_1|$ of these sets of the form $C_i \cap C_1$, and therefore $m - 1 \le n - |C_1| \le n - 1 \implies m \le n$, as desired.
- Now suppose that $|C_j| > t$ for all $1 \le j \le m$. Here's where the linear algebra comes in: we again construct incidence vectors $v_1, \cdots, v_m \in \{0, 1\}^n$, but this time because we don't have parity restrictions we will consider these vectors living in $\mathbb{R}^n$ instead. Then we find that $v_i \cdot v_j = t$ for $i \ne j$ and $v_i \cdot v_i > t$ for all i, meaning that if we form the $n \times m$ matrix A with columns $v_1, \cdots, v_m$, then $A^T A$ has off-diagonal entries t and diagonal entries $|C_i| > t$. Thus, we have
$$A^T A = tJ + D,$$
where J is the all-ones matrix and D is a diagonal matrix with strictly positive entries. But again $\text{rank}(A^T A) \le \text{rank}(A) \le n$, so it suffices to show that our matrix $A^T A$ has full rank. Indeed, this is because our matrix is

positive definite (in other words, $x^T(A^TA)x > 0$ for all nonzero $x \in \mathbb{R}^m$), since tJ is positive semidefinite and D is positive definite; more explicitly, if x is the column vector with entries $(x_1, \cdots, x_m)$, then (viewing expressions of the form x^TMx as bilinear form calculations)

$$x^T(A^TA)x = x^T(tJ)x + x^TDx = t\sum_{i,j=1}^{m} x_ix_j + \sum_{i=1}^{m}(|C_i| - t)x_i^2.$$

But these terms simplify because the first term becomes $t(x_1 + \cdots + x_m)^2 \geq 0$, and the second term is a sum of strictly positive terms unless the corresponding x_is are zero. So the only way for $x^T(A^TA)x \leq 0$ is if $x_i = 0$ for all i, as desired. So A^TA has full rank m (otherwise we would have a nonzero vector x in the kernel, which would have implied that $x^T(A^TA)x = 0$), and thus $m \leq n$, completing the proof.

□

3 February 8, 2022

Last week, we discussed theorems about families of sets with certain rules about their intersections. We'll discuss another result of this type today, but we'll now consider two different families of sets together:

Theorem 10 (Skew Bollobás' Theorem)
Suppose $A_1, \cdots, A_n$ are sets of size $|A_i| = r$, and $B_1, \cdots, B_n$ are sets of size $|B_i| = s$ (with no restriction on the ground set that they live in). Suppose that $A_i \cap B_i = \varnothing$ for all $1 \leq i \leq n$, but $A_i \cap B_j \neq \varnothing$ for $1 \leq i < j \leq n$ (note that this is not the same as $i \neq j$). Then $n \leq \binom{r+s}{s}$.

This is pretty different from the other results we've been seeing, specifically because there's no restriction on what elements our A_is and B_is can contain, and yet we still have an upper bound on n.

First, we show that this bound is actually sharp: let $n = \binom{r+s}{s} = \binom{r+s}{r}$, and let the A_is be the set of all r-element subsets of $\{1, 2, \cdots, r+s\}$. If we then define $B_i = \{1, 2, \cdots, r+s\} \setminus A_i$, we know that $A_i \cap B_i = \varnothing$, and we also have $A_i \cap B_j \neq \varnothing$ because there are only $r+s$ total elements (so a set of size r and a set of size s can only be disjoint if they are complements, and $A_i \neq A_j$).

Fact 11
Notice that our theorem statement only required $A_i \cap B_j \neq \varnothing$ for $i < j$, but in fact this construction achieves the condition for $i > j$ as well. Bollobás only proved this theorem for the case where $i \neq j$, and that result turns out to be significantly easier – this "skew" version with $i < j$ is what we'll show now.

Before we do the actual proof, we'll begin by thinking about the **moment curve** in $\mathbb{R}^d$, which is the set

$$\left\{ \begin{bmatrix} 1 \\ x \\ \vdots \\ x^{d-1} \end{bmatrix} : x \in \mathbb{R} \right\} \subseteq \mathbb{R}^d.$$

Lemma 12
Any d vectors on this moment curve are linearly independent.

Proof. Indeed, if we have our vectors $\begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_1^{d-1} \end{bmatrix}, \begin{bmatrix} 1 \\ x_2 \\ \vdots \\ x_2^{d-1} \end{bmatrix}, \cdots, \begin{bmatrix} 1 \\ x_d \\ \vdots \\ x_d^{d-1} \end{bmatrix}$, we can consider the determinant of the matrix formed by putting them together,

$$\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_d^{d-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{d-1} & x_2^{d-1} & \cdots & x_d^{d-1} \end{bmatrix}.$$

This is the **Vandermonde determinant**, and it turns out this expression is equal to $\prod_{i<j}(x_j - x_i)$ (by considering the degree of the polynomial in the x_is and using the factor theorem). And because we assumed that we had distinct x_is, this determinant is indeed nonzero.

But if we're not satisfied with that proof, here's another way to think about it: suppose this determinant is zero, so that the **rows** are linearly dependent as well. Thus, there exist constants $a_0, \cdots, a_{d-1}$ so that

$$0 = a_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + a_1 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} + \cdots + a_{d-1} \begin{bmatrix} x_1^{d-1} \\ x_2^{d-1} \\ \vdots \\ x_d^{d-1} \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x_1 + \cdots + a_{d-1} x_1^{d-1} \\ a_0 + a_1 x_2 + \cdots + a_{d-1} x_2^{d-1} \\ \vdots \\ a_0 + a_1 x_d + \cdots + a_{d-1} x_d^{d-1} \end{bmatrix}.$$

In other words, $x_1, \cdots, x_d$ are all roots of the degree at-most-$(d-1)$ polynomial $a_0 + a_1 x + \cdots + a_{d-1}x^{d-1}$, and because the x_is are distinct this can only occur if the polynomial is the zero polynomial. □

We'll also need some additional linear algebra tools to solve this problem, namely the concept of an **exterior algebra** (which might take some time to get used to):

Definition 13

Let W be a vector space over $\mathbb{R}$ (this also works over other fields of characteristic not equal to 2). The ***k*th exterior power of *W***, denoted $\Lambda^k(W)$, is the quotient of the tensor product $W \otimes W \otimes \cdots \otimes W$ (k times) by relations of the form

$$z_1 \otimes \cdots \otimes z_{j-1} \otimes x \otimes y \otimes z_{j+1} \otimes \cdots \otimes z_k = -z_1 \otimes \cdots \otimes z_{j-1} \otimes y \otimes x \otimes z_{j+1} \otimes \cdots \otimes z_k,$$

for all $1 \le j \le k-1$ and $x, y, z_i \in W$.

In other words, we start with the tensor product $W^{\otimes k}$ – luckily, because we're tensoring over a field rather than a ring, things are pretty simple. Specifically, we can write down a basis of $W^{\otimes k}$ by fixing a basis of W and then looking at "pure tensors" of the form $z_1 \otimes z_2 \otimes \cdots \otimes z_k$, where each z_i is a basis element of W. Then we can take arbitrary linear combinations of such pure tensors, and that gives us all of the elements of $W^{\otimes k}$.

From there, the exterior algebra is formed by saying that if we flip the order of two adjacent elements, then we get the same element back but with a negative sign.

Fact 14

Let $w_1, \cdots, w_\ell$ be a basis of W. Then elements of the form

$$w_{j_1} \wedge w_{j_2} \wedge \cdots \wedge w_{j_k}, 1 \le j_1 < j_2 < \cdots < j_k \le \ell,$$

which are the images of $w_{j_1} \otimes w_{j_2} \otimes \cdots \otimes w_{j_k}$ after quotienting, form a basis of $\Lambda^k(W)$.

(We ask that the j_i indices are strictly increasing to avoid linear dependence due to switching – in particular, if two indices are the same, then swapping until we get those indices next to each other shows us that the element is its own negative by swapping again, meaning that we just have zero.) Because of this fact, we know that $\dim \Lambda^k(W) = \binom{\dim W}{k}$, and thus if $k > \dim W$ the exterior power is just the zero vector space. More generally, we have the following fact:

Fact 15

For any $w_1, \cdots, w_k \in W$, we have $w_1 \wedge w_2 \cdots \wedge w_k \neq 0$ if and only if the w_is are linearly independent in W.

With that, we're ready to return to the original proof:

Proof of Theorem 10. Suppose we have our sets $A_1, \cdots, A_n, B_1, \cdots, B_n$ – these are all subsets of some (potentially large but finite) set of size M, which we'll label $\{1, 2, \cdots, M\}$. (Even if we had an infinite set of A_is and B_is, this proof argument would still allow us to follow this argument by truncating those infinite sets at, say, $\binom{r+s}{s} + 1$ elements.) Let $w_1, \cdots, w_M$ be M distinct vectors on the **moment curve** in $\mathbb{R}^{r+s}$; by Lemma 12, any set of $(r+s)$ of the vectors $w_1, \cdots, w_M$ will be linearly independent. (This linear independence for **any** set of $(r+s)$ vectors is actually all we need from the moment curve.) We now define

$$a_i = w_{j_{i,1}} \wedge \cdots \wedge w_{j_{i,r}} \in \Lambda^r(\mathbb{R}^{r+s}), \text{ where } j_{i,1} < \cdots < j_{i,r} \text{ are elements of } A_i$$

in other words, we wedge the elements of A_i together in increasing order (so if $A_1 = \{2, 3, 5\}$, then $a_1 = w_2 \wedge w_3 \wedge w_5$). Similarly, we also define

$$b_i = w_{j'_{i,1}} \wedge \cdots \wedge w_{j'_{i,s}} \in \Lambda^r(\mathbb{R}^{r+s}), \text{ where } j'_{i,1} < \cdots < j'_{i,s} \text{ are elements of } B_i.$$

But now we can form $a_i \wedge b_j \in \Lambda^{r+s}(\mathbb{R}^{r+s})$ for any $1 \le i, j \le M$ (because a_i is always a wedge of r elements, and b_i is always a wedge of s elements). Then $a_i \wedge b_i \neq 0$ for any i, because we're wedging together $(r+s)$ vectors from the moment curve, which are linearly independent and then we use Fact 15. On the other hand, $a_i \wedge b_j = 0$ for any $i < j$ because of our repeated element (which means we don't have linear independence). This is the point where it's important that our w_i vectors live in $\mathbb{R}^{r+s}$!

So now notice that $a_1, \cdots, a_n \in \Lambda^r(\mathbb{R}^{r+s})$ are all vectors in the $\binom{r+s}{s}$-dimensional vector space, so to show that $n \le \binom{r+s}{s}$, it suffices to show that the a_is are linearly independent. Indeed, suppose we have some linear combination of the form

$$\lambda_1 a_1 + \cdots + \lambda_n a_n = 0$$

in $\Lambda^r(\mathbb{R}^{r+s})$ for some $\lambda_i \in \mathbb{R}$. We can now wedge with $b_n, b_{n-1}, \cdots, b_1$ in that order to show that $\lambda_n, \lambda_{n-1}, \cdots \lambda_1$ are zero, but more concisely we can let $j \in \{1, \cdots, n\}$ be the maximum index with $\lambda_j \neq 0$. If wedge our equation above with b_j, we then find that

$$(\lambda_1 a_1 + \cdots + \lambda_n a_n) \wedge b_j = 0.$$

We can now distribute (because tensoring is multilinear, so is wedging), so that this equation becomes

$$\lambda_1 a_1 \wedge b_j + \cdots + \lambda_n a_n \wedge b_j = 0.$$

But now for $i < j$, $a_i \wedge b_j = 0$, and for $i > j$, $\lambda_i = 0$. So the only term that remains is $\lambda_j a_j \wedge b_j$, and thus (because $a_j \wedge b_j \neq 0$) $\lambda_j = 0$, a contradiction. Thus no linear combination exists, the a_is are linearly independent, and $n \le \binom{r+s}{r}$. □

Remark 16. *On our homework, we'll see that we can also do a two-family version of the odd-town theorem, and we'll see how probabilistic considerations show up there. But for this skew Bollobás result, only linear algebra proofs are currently known (though there are others which only use the moment curve or only use the exterior algebra).*

Fact 17

Last week, we tried flipping the roles of "even" and "odd" in the even- and odd-town theorems, but that won't give us anything interesting for the skew Bollobaś result – in all three other modifications of the theorem conditions, we can easily have infinitely many sets. So that's another reason that we should see this result to be surprising!

4 February 10, 2022

We'll move on to a different question (of a geometric flavor) today, starting with a warmup version:

Problem 18

What is the maximum number of lines in $\mathbb{R}^2$ through the origin, such that the angle between any two of them is the same?

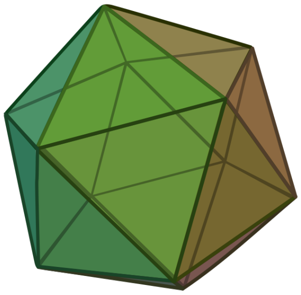
If we have two lines at an angle of α apart, then adding a third line must require that line to be at an angle of α from both of the first two lines. This can be achieved by having three lines at 60 degree angles from each other, and indeed we can check that $\boxed{3}$ is the maximum. (The reason that we ask the lines to be through the origin is to avoid infinitely many parallel lines.) We'll now move up by one dimension:

Problem 19

What is the maximum number of lines in $\mathbb{R}^3$ through the origin, such that the angle between any two of them is the same?

Two ways that we can again get **3** lines through the origin is to consider three pairwise orthogonal lines, or to use the construction from $\mathbb{R}^2$. And one way we can get **4** lines through the origin are to take the vertices of a tetrahedron centered at the origin and connect them to the center (or equivalently to do the same thing with a cube – that's the same construction but potentially easier to visualize).

It is then natural to try using other Platonic solids to get a similar symmetry, and indeed we can do this by taking the **longest diagonals of an icosahedron**:



Since icosahedrons have 12 vertices, and connecting opposite vertices passes through the center, this gives us $\boxed{6}$ lines. And indeed, this gives us equiangular lines because to get from any diagonal to another, we look at one endpoint, take one of the five adjacent vertices, and form the diagonal from that vertex. And this is in fact the best that we can do:

Theorem 20

There are at most $\binom{d+1}{2}$ lines through the origin in $\mathbb{R}^d$ such that the angle between any two of them is the same.

Proof. Suppose we have a set of equiangular lines $\ell_1, \cdots, \ell_n$, all separated by some angle $\phi \in (0, \frac{\pi}{2}]$. We aim to prove that $n \le \binom{d+1}{2}$, and we'll do so using linear algebra. For all $1 \le i \le n$, let v_i be a unit vector in the direction of ℓ_i – there are always two choices, but we can arbitrarily pick between them. Then based on our restrictions, we have

$$v_i \cdot v_j = \begin{cases} 1 & i = j, \\ \pm \cos \phi & i \ne j, \end{cases}$$

because the dot product is the length of the projection of v_j onto v_i, where the ± sign depends on the direction of the line that we've chosen in each case. This time, the strategy is not to show linear independence of the vectors v_i (that won't work) – instead, for each $1 \le i \le n$, we consider the $d \times d$ matrix $v_i v_i^T$. This is symmetric because $(v_i^T v_i)^T = (v_i)^T (v_i^T)^T = v_i^T v_i$, and we can notice that the dimension of the space of symmetric $d \times d$ matrices is $d + (d-1) + \cdots + 1 = \frac{d(d+1)}{2} = \binom{d+1}{2}$. Thus, it suffices to prove **linear independence of the matrices** $v_1 v_1^T, v_2 v_2^T, \cdots, v_n v_n^T$.

Indeed, suppose we have some linear combination of those matrices satisfying

$$\sum_{i=1}^{n} \lambda_i v_i v_i^T = 0$$

(note that this is an equality of $d \times d$ matrices). To turn this into an equality of numbers, multiply by v_j^T from the left and v_j from the right: since $v_i^T v_j = v_i \cdot v_j$, we in fact have

$$\sum_{i=1}^{n} \lambda_i v_j^T v_i v_i^T v_j = v_j^T 0 v_j \implies \sum_{i=1}^{n} \lambda_i (v_i \cdot v_j)^2 = 0.$$

But now the ± ambiguity in the dot products goes away, and we're left with the relation

$$\lambda_j + \cos^2 \phi \sum_{i \ne j} \lambda_i = 0$$

for any j. In particular, this means that

$$\cos^2 \phi \sum_{i=1}^{n} \lambda_i = (\cos^2 \phi - 1)\lambda_j,$$

and because $0 < \phi \leq \frac{\pi}{2}$ this can only hold if all λ_js are equal. Since the left and right sides of this equation will always have opposite sign, this will only occur if λ_i are all zero. Thus linear independence is shown, and $n \leq \binom{d+1}{2}$. □

This bound is tight for $d = 2, 3$, but it is not tight in general – nevertheless, we do have a lower bound which is also quadratic in d. So the true answer is known up to a constant factor, which is often the best we can do in extermal combinatorics.

Fact 21

On the other hand, this flavor of question can be very difficult – those lower bound constructions all have common angle tending to 90 degrees, and if we instead ask the related question of "what is the maximum number of lines in $\mathbb{R}^d$ with some fixed common angle α?", the answer is instead at most linear in d, and refining those bounds is ongoing research (partially done by some of the students in this room!)

Notice that our n lines $v_1, \cdots, v_n$ give us n points on the unit sphere, and those points have the special property that there are only two possible distances between them (depending on the orientation of the lines). That motivates the next question of **point-sets with only two distances**, and we'll again start off with a preliminary question:

Problem 22

What is the maximum number of points in $\mathbb{R}^d$ such that any two points are the same distance apart?

The answer is $\boxed{d+1}$, formed by taking a regular d-dimensional simplex – finding a nice way to prove this (such as with linear algebra similar to Fisher's inequality) is on our homework. So such a situation is extremely constrained, and we'll relax the constraints to get an interesting problem now:

Problem 23

What is the maximum number of points in $\mathbb{R}^d$ such that any two points are one of two distances apart?

Notice that we are allowing any two distances in this problem, while in the equiangular lines case the two distances were actually related to each other by geometric considerations. But it turns out that the answer even with this extra freedom is the same to leading order:

Theorem 24

Let $p_1, \cdots, p_n \in \mathbb{R}^d$ be distinct points such that there are only two distinct distances between them. Then $n \leq \frac{1}{2}(d^2 + 5d + 4)$.

In this proof, we'll see the introduction of a cool technique which will come up more in future lectures as well.

Start of proof. Let the two distances between the points be $a, b \neq 0$ (if there's only one distance then we can let b be some arbitrary positive number). We'll be showing linear independence again, but this time we need to be a bit more clever with the objects for which we are proving that linear independence.

Specifically, for all $1 \leq j \leq n$, consider the function $f_j : \mathbb{R}^d \to \mathbb{R}$ given by

$$f_j(x) = (||x - p_j||^2 - a^2)(||x - p_j||^2 - b^2).$$

(Here, $||x - p_j||$ denotes the distance between the points x and p_j.) This is a real-valued function satisfying $\boxed{f_j(p_i) = 0 \text{ for all } i \neq j}$ (since the distance between p_i and p_j is either a or b) and $\boxed{f_j(p_j) = a^2 b^2}$.

We now claim that the functions $f_1, \cdots, f_n$ are linearly independent. Indeed, suppose we have $\lambda_1 f_1 + \cdots + \lambda_n f_n = 0$ (this is an equation of functions); evaluating both sides at p_i makes most terms disappear by the boxed relations above, so we're just left with $\lambda_i f_i(p_i) = \lambda_i a^2 b^2 = 0$, so that $\lambda_i = 0$. Repeating this for all i shows the claim.

The space of functions from $\mathbb{R}^d$ to $\mathbb{R}$ is infinite-dimensional, so the dimension considerations we've been doing so far are not directly useful for us yet. But now we can take a closer look at $f_j(x)$ and understand why we chose functions of that form – notice that the f_js are all **polynomials** in the d coordinates $x = (x_1, \cdots, x_d)$, because $||x - p_j||^2 = \sum_{i=1}^d (x_i - p_j^{(i)})^2$ expands out to a sum of squares in those coordinates. So all f_js are in the (much smaller) space of polynomials in d coordinates – that space is still infinite-dimensional, but next lecture we'll continue to narrow down some properties of f_j (for example, they all have degree at most 4, finally giving us a finite-dimensional vector space) and finish the proof. □

5 February 15, 2022

Last lecture, we considered sets of points where there are only two distinct distances among the points, and we started proving that there could only be at most $\frac{1}{2}(d^2 + 5d + 4)$ in $\mathbb{R}^d$ (quadratic in d, while when we only get one distinct distance we get only linear in d). We'll continue the proof here:

Proof, continued. Recall the main steps from last time: for each point p_i in our set, we defined a function

$$f_i(x) = \left(||x - p_i||^2 - a^2\right)\left(||x - p_i||^2 - b^2\right),$$

and then we showed that the different functions $f_1, \cdots, f_n$ are linearly independent. (because $f_i(p_j) = 0$ except when $i = j$). We mentioned last time that the space of all functions $\mathbb{R}^d \to \mathbb{R}$ is infinite-dimensional, so linear independence doesn't seem to help yet. But the next step is to reduce the size of the subspace under consideration – we saw that because we're taking squared norms of distances, f_i is actually always a polynomial, and it has degree at most 4. Since that space has dimension $\binom{d+4}{4}$, we get a quartic bound for n, and it remains to narrow down the bound further.

To get a more refined bound now, we'll expand out the squared norms:

$$f_i(x) = \left(||x||^2 - 2p_i \cdot x + ||p_i||^2 - a^2\right)\left(||x||^2 - 2p_i \cdot x + ||p_i||^2 - b^2\right).$$

We could expand out all 16 terms, but we can look at the terms in order of degree of $x = (x_1, \cdots, x_d)$:

$$= ||x||^4 - 4(p_i \cdot x)||x||^2 + (\text{some degree in } (x_1, \cdots, x_d) \text{ of degree} \le 2).$$

Here, we should remember that $||x||^2 = x_1^2 + \cdots + x_d^2$ is indeed quadratic in the x_is. Specifically, this polynomial is actually

$$f_i(x) = (x_1^2 + \cdots + x_d^2)^2 - 4\sum_{i=1}^n p_i^{(k)} x_k (x_1^2 + \cdots + x_d^2) + (\text{some degree in } (x_1, \cdots, x_d) \text{ of degree} \le 2),$$

where $p_i^{(k)}$ is the kth coordinate of the point p_i. Thus, $f_1, \cdots, f_n$ live in the span of the following polynomials – we have to take all monomials of degree at most 2 to get the unspecified degree-2 part, and then we need to account for the linear combinations that we can get from the leading terms:

$$\left\{1,\ x_j,\ x_j^2,\ x_i x_j,\ (x_1^2 + \cdots + x_d^2)^2,\ x_j(x_1^2 + \cdots + x_d^2) : 1 \le i, j \le d, i \ne j\right\}.$$

That means $f_1, \cdots, f_n$ are linearly independent elements of a vector space of dimension at most (just adding up how many elements we have in the spanning list)

$$1 + d + d + \binom{d}{2} + 1 + d = \frac{1}{2}(d^2 + 5d + 4),$$

and thus $n \le \frac{1}{2}(d^2 + 5d + 4)$, as desired. □

Remark 25. *It turns out we can improve on this proof, because we forgot about the information in the degree-2-polynomial part: it turns out we can have* $n \le \binom{d+2}{2} = \frac{1}{2}(d^2 + 3d + 2)$*, and that's going to be a homework assignment for us.*

It's natural now to ask about lower bounds (in particular trying to construct such a point-set): one method is to take the set of vectors in the Boolean cube $\{0, 1\}^d$ with exactly two ones. Then $n = \binom{d}{2} = \frac{1}{2}(d^2 - d)$, and the distances between the points are either 2 (if there are no overlaps between which indices are 1s) or $\sqrt{2}$ (if there is one overlap).

We can further improve on this by noticing that the construction above actually has all points within a hyperplane $x_1 + \cdots + x_d = 2$, so it essentially lives in $\mathbb{R}^{d-1}$ instead. So we can replace $n \le \binom{d}{2}$ with $n \le \binom{d+1}{2} = \frac{1}{2}(d^2 + d)$ by doing a "back-construction," doing this in $\mathbb{R}^{d+1}$ and then looking at the **affine hyperplane** ("affine" meaning that the hyperplane doesn't need to go through the origin) isomorphic to $\mathbb{R}^d$ containing all of our points. (Equivalently, we can consider a simplex of $(d + 1)$ points and connect midpoints of edges – then the only two cases are if the edges are adjacent – equilateral triangle – or not – tetrahedron.) And that's (probably) the best-known construction so far.

That's all we'll say about point sets of two distances for now – the key point was to do linear independence on certain smartly-constructed functions. And we'll use this polynomial technique to go back to the skew Bollobás theorem and prove it in another way not involving exterior algebras. As a reminder, the statement is as follows: suppose $A_1, \cdots, A_n$ are sets of size $|A_i| = r$, and $B_1, \cdots, B_n$ are sets of size $|B_i| = s$ (with no restriction on the ground set). Further suppose that $A_i \cap B_i = \varnothing$ for all $1 \le i \le n$, but $A_i \cap B_j \ne \varnothing$ for $1 \le i < j \le n$. Then $n \le \binom{r+s}{s}$.

Alternative proof of skew Bollobás. Like in Lecture 3, we'll need our lemma about the moment curve, which says that any d distinct points $(1, x, \cdots, x^{d-1})$ in $\mathbb{R}^d$ are linearly independent (by computing the Vandermonde determinant). Fix notation so that $A_1, \cdots, A_n, B_1, \cdots, B_n$ are subsets of $\{1, \cdots, M\}$ for some positive finite M, and pick M distinct points $p_1, \cdots, p_M$ on the moment curve in $\mathbb{R}^{r+1}$ (note that this is different from picking points from the moment curve in $\mathbb{R}^{r+s}$ like we did last time). Then any $(r + 1)$ points among $p_1, \cdots, p_M$ are linearly independent.

The functions we write down this time are slightly more annoying: if $B \subseteq \{1, \cdots, M\}$ is an arbitrary s-element subset (such as $B_1, \cdots, B_n$, but we want to simplify notation), we define $f_B : \mathbb{R}^{r+1} \to \mathbb{R}$ via

$$f_B(y) = \prod_{i \in B} (p_i \cdot y).$$

This is a homogeneous polynomial of degree s in $y = (y_1, \cdots, y_{r+1})$ (because each term $p_i \cdot y$ is a linear combination of the y_is, and there are $|B| = s$ of them). Thus, f_B is always in the space of homogeneous polynomials in $\mathbb{R}[y_1, \cdots, y_{r+1}]$ of degree s, which has dimension $\binom{r+s}{r}$ by stars and bars. So now if we can show that $f_{B_1}, \cdots, f_{B_n}$ are linearly independent, we'll have the desired bound $n \le \binom{r+s}{r}$.

To show linear independence, recall that with the point-set proof, we plugged in particular points into our polynomials such that $f_i(p_j) = 0$. Similarly, we want to ask here when $f_B(y) = 0$, which occurs if and only if y is orthogonal to p_i for some $i \in B$. Unfortunately, the way we set this up, p_i and p_j are likely not orthogonal, so we have to be more careful. For each $1 \le i \le n$, associate a vector q_i with the set A_i as follows: because A_i has r elements within $\{1, \cdots, M\}$, that gives us r linearly independent points p_k in $\mathbb{R}^{r+1}$ among $p_1, \cdots, p_M$, and now q_i can be **some vector**

orthogonal to all r of those points (in other words, take the r-dimensional hyperplane formed by the points p_k, and take a vector orthogonal to that plane). We can then notice that

$$q_i \cdot p_\ell = \begin{cases} 0 & \ell \in A_i, \\ \neq 0 & \ell \notin A_i, \end{cases}$$

because if $\ell \in A_i$, then p_ℓ is one of the points in the hyperplane that q_i must be orthogonal to, and **if** (for the sake of contradiction) $\ell \notin A_i$ but $q_i \cdot p_\ell = 0$, then p_ℓ would be in the hyperplane spanned by the p_ks for $k \in A_i$, and that would be bad because it would be $(r+1)$ points on the moment curve that are linearly dependent. So now we have the orthogonality relations that are necessary for continuing on with the proof: evaluating our functions at the points q_i, we have

$$f_{B_j}(q_i) = 0 \iff q_i \cdot p_\ell = 0 \text{ for some } \ell \in B_j \iff \ell \in A_i \text{ for some } \ell \in B_j,$$

so $f_{B_j}(q_i) = 0$ **if and only if** $A_i \cap B_j \neq \varnothing$. We can now finally use the properties of A_i and B_j from the statement of skew Bollobás:

$$f_{B_j}(q_i) = \begin{cases} \neq 0 & i = j, \\ 0 & i < j, \\ \text{(unknown)} & i > j, \end{cases}$$

and just like last lecture this is enough to show linear independence. After all, if we have some nonzero linear combination $\lambda_1 f_{B_1} + \cdots + \lambda_n f_{B_n} = 0$ (this is an equality of functions $\mathbb{R}^{d+1} \to \mathbb{R}$), plugging in the smallest i such that $\lambda_i = 0$, we have $\lambda_1 = \cdots = \lambda_{i-1} = 0$, and also $f_{B_{i+1}}(q_i) = \cdots = f_{B_n}(q_i) = 0$. So all the terms go away except $\lambda_i f_{B_i}(q_i)$, which is nonzero because we assumed $\lambda_i \neq 0$ and we know $f_{B_i}(q_i) \neq 0$, which is a contradiction. Thus linear independence is shown, and $n \leq \binom{r+s}{r}$, the dimension of the space we've been considering for the f_Bs. □

6 February 17, 2022

We'll discuss a result by Frankl and Wilson, titled "medium-size intersection is hard to avoid." The result is interesting on its own, but we'll also use it later to find a counterexample to a geometric conjecture. Much like other results we've seen in this class, this is a statement about sets and their intersections:

Theorem 26 (Frankl, Wilson (special case))
Let p be a prime, and let $\mathcal{F}$ be a family of subsets of $\{1, \cdots, n\}$ each of size $(2p-1)$. If no two members of $\mathcal{F}$ intersect in exactly $(p-1)$ elements, then $|\mathcal{F}| \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1}$.

Notice that the total number of subsets of $\{1, \cdots, n\}$ of size $2p-1$ is $\binom{n}{2p-1}$, which is much larger than $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1}$ for large n (because the former is a polynomial of degree $(2p-1)$, while the latter is a polynomial of degree $(p-1)$). We can also get a lower-bound construction to see that this bound is asymptotically tight (for large n and fixed p): we can get $|\mathcal{F}| = \binom{n-p}{p-1}$ by taking $\mathcal{F}$ to be the set of all subsets of $\{1, \cdots, n\}$ that contain $\{1, \cdots, p\}$. (Then the intersection of any two subsets has size at least $p > p-1$.)

Proof. We'll again do a linear independence proof based on cleverly constructed functions. For each set $A \in \mathcal{F}$, we once again define its characteristic vector $v_A \in \{0,1\}^n$ (1 in the ith index if A contains i), and we will also define a function f_A which takes vectors x in $\{0,1\}^n$ as inputs. Since our theorem statement has to do with intersections between sets, we want to sum up the entries of x in the indices of A (equivalently, do our usual dot product argument), and furthermore we want to be able to make linear independence arguments using the fact that $f_A(v_B) = 0$ for $A \neq B$ (like we've previously done). This seems to motivate the definition

$$f_A(x) \overset{?}{=} \prod_{s \in \{0, \cdots, 2p-1\}, s \neq p-1} \left[\left(\sum_{i \in A} x_i \right) - s \right],$$

but the degree of such a polynomial turns out to be too large for us to get the bound that we want. So instead, we'll only multiply over the smaller values of s:

$$\boxed{f_A(x) = \prod_{s=0}^{p-2} \left[\left(\sum_{i \in A} x_i \right) - s \right]},$$

and to compensate for the fact that $f_A(v_B)$ doesn't seem to vanish if the intersection between A and B is too large, **we'll have f_A take values in $\mathbb{F}_p$.**

Remark 27. *Notice that it's equivalent over $\mathbb{F}_p$ to define this function as*

$$f_A(x) = \left[\left(\sum_{i \in A} x_i \right) - (p-1) \right]^{p-1} - 1,$$

basically by using Fermat's little theorem.

We can indeed see that $f_A(x)$ is nonzero if and only if all of the factors in the product are all nonzero in $\mathbb{F}_p$ – since the product runs over all residues mod p except one of them, the only way for $f_A(x) \neq 0$ is if $\sum_{i \in A} x_i = p - 1$ in $\mathbb{F}_p$. Thus, $\boxed{f_A(v_B) \neq 0}$ if and only if $\sum_{i \in A} (v_B)_i = p - 1$, if and only if $|A \cap B| = p - 1$ in $\mathbb{F}_p$. And by the theorem statement and the fact that $|A \cap A| = 2p - 1 = p - 1$ in $\mathbb{F}_p$, this indeed occurs **if and only if** $\boxed{A = B}$.

Now, the functions $\{f_A(x) : A \in \mathcal{F}\}$ are linearly independent for the same reason as usual: for any linear combination $\sum_A \lambda_A f_A = 0$ (as an equality of functions $\{0,1\}^n \to \mathbb{F}_p$), if we evaluate both sides at v_B, we're left with $\lambda_B f_B(v_B) = 0$, so that $\lambda_B = 0$ for all B. So now it suffices to calculate the dimension of the space of functions that the f_As live in – the dimension of the space of functions $\{0,1\}^n \to \mathbb{F}_p$ is 2^n, but $|\mathcal{F}| \leq 2^n$ is a useless bound.

Instead, notice that all f_As are polynomials, and their degrees are all $p - 1$, so we can consider the subspace of all polynomials of degree at most $p - 1$. This space looks like it has dimension $\binom{n+p-1}{p-1}$ by stars and bars, which is not quite good enough. But **there's a subtlety here**; remember that our domain is $\{0,1\}^n$, so that x_i^2 is actually the same polynomial as x_i. So powers of 2 always disappear, and over the space we're considering, the dimension of the subspace is indeed just $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{p-1}$ (because for each degree d, we pick d different elements of $\{x_1, \cdots, x_n\}$ to multiply together to form a monomial). And we can take our functions f_A, multiply them out, and then remove the higher exponents from the various x_is – this doesn't change the function and keeps the degree at most $(p-1)$. Thus $|\mathcal{F}|$ must indeed be smaller than this dimension, proving the result. □

(We can ask about generalizing this statement to non-prime p or more general sizes of sets and intersections – we might talk more about this next lecture.) With that, we'll now turn to the geometric application, starting with a somewhat unmotivated lemma that will be used as a blackbox:

Corollary 28

Let p be a prime, and let $n = 4p$. If $\mathcal{F}$ is a family of $(2p-1)$-element subsets of $\{1, \cdots, n\} = \{1, \cdots, 4p\}$, such that no two members of $\mathcal{F}$ intersect in exactly $(p-1)$ elements. Then

$$\frac{|\mathcal{F}|}{\binom{4p}{2p-1}} \le 1.1^{-4p}.$$

In other words, we get an exponentially small bound on the fraction of total $(2p-1)$-element subsets we can have at once.

Proof. By Theorem 26, we know that

$$\begin{aligned} |\mathcal{F}| &\le \binom{4p}{p-1} + \binom{4p}{p-2} + \cdots + \binom{4p}{0} \\ &\le \binom{4p}{p}\left[\frac{1}{3} + \left(\frac{1}{3}\right)^2 + \cdots + \left(\frac{1}{3}\right)^p\right], \end{aligned}$$

because inductively, going from $\binom{n}{k}$ to $\binom{n}{k-1}$ requires multiplying by $\frac{k}{n-k+1}$, which is always at most $\frac{1}{3}$ because we always have $n \ge 4k$ here. Evaluating the geometric series, we find that

$$|\mathcal{F}| \le \frac{1}{2}\binom{4p}{p} \implies \frac{|\mathcal{F}|}{\binom{4p}{2p-1}} \le \frac{\frac{1}{2}\binom{4p}{p}}{\binom{4p}{2p-1}} = \frac{1}{2}\frac{(2p-1)!(2p+1)!}{(3p)!p!}.$$

Expanding this out and canceling factors, we find that

$$\frac{|\mathcal{F}|}{\binom{4p}{2p-1}} \le \frac{(2p-1)(2p-2)\cdots(p+1)}{(3p(3p-1)\cdots(2p+2)} \le \frac{1}{2}\left(\frac{2}{3}\right)^{p-1}$$

because each corresponding term in the numerator and denominator have ratio at most $\frac{2}{3}$. Thus we indeed find

$$\frac{|\mathcal{F}|}{\binom{4p}{2p-1}} \le \left(\frac{2}{3}\right)^p \le 1.1^{-4p}.$$

□

We can now present the geometric conjecture that we've been alluding to:

Definition 29

The **diameter** of a set $X \subseteq \mathbb{R}^d$, denoted $\text{diam}(X)$, is given by

$$\text{diam}(X) = \sup_{x,y \in X} ||x-y||.$$

Conjecture 30 (Borsuk's conjecture)

Every subset $X \in \mathbb{R}^d$ with finite diameter can be partitioned into $(d+1)$ subsets $X_1, \cdots, X_{d+1}$ such that $\text{diam}(X_i) < \text{diam}(X)$ for all i.

This conjecture is clearly false if we use fewer than $d+1$ subsets, because we can take a simplex of $(d+1)$ vertices of side length 1 – then the diameter of the set is 1, but the pigeonhole principle tells us that if we partition into d

sets, one of those must have at least two points, and thus the diameter of that set would be 1. For historical context, Borsuk first considered the unit ball, in which he proved that partitioning it into d subsets of smaller diameter was impossible but that $(d+1)$ subsets was possible. And this conjecture was not directly by Borsuk, but it was motivated by his result and quite widely believed (it was proved for dimension $d = 2, 3$ and also for **smooth** convex sets in $\mathbb{R}^d$ – note that the convexity is not the issue, because the diameter of a set is the same as the diameter of its convex hull). But here's how the conjecture was disproved (in fact, shown to be very false):

Theorem 31 (Kahn and Kalai (1993))
Let p be a prime and $n = 4p$. Then there exists a set $X \subseteq \mathbb{R}^{n^2}$ of finitely many points (so finite diameter), such that for every partition of X into subsets $X_1, \cdots, X_m$ for $m < 1.1^n$, we have $\text{diam}(X_j) = \text{diam}(X)$ for some X.

This result says that we can take exponentially many subsets (with exponent proportional to the square root of the dimension of the space) and still not be able to reduce the diameter! In particular, whenever $1.1^n > n^2 + 1$ (for example, $n \geq 100$ and n is four times a prime), we get a counterexample to Borsuk's conjecture. On the other hand, we can show that we can always reduce the diameter using exponentially many subsets (with a larger base), but we'll talk about that next time as well.

We'll do the preparatory steps for the proof and show the result next time. As a spoiler, we'll be looking at all $(2p-1)$-element subset of $\{1, \cdots, 4p\}$ and associate it to a point in X. What's interesting is that we're in dimension n^2 here, and we'll do that with **tensor product notation**: for $x, y \in \mathbb{R}^n$, we let $x \otimes y \in \mathbb{R}^{n^2}$ have entries $x_i y_j$ for $1 \leq i, j \leq n$ (in other words, consider the entries of the $n \times n$ matrix xy^T, but we **think of this as a vector** in $\mathbb{R}^{n^2}$).

The claim is now that for any $x, y \in \mathbb{R}^n$, we have

$$(x \otimes x) \cdot (y \otimes y) = (x \cdot y)^2,$$

where the dot products are in $\mathbb{R}^{n^2}$ and $\mathbb{R}^n$ on the left and right sides, respectively. Indeed, this is just a computation:

$$(x \otimes x) \cdot (y \otimes y) = \sum_{i=1}^n \sum_{j=1}^n (x \otimes x)_{ij} (y \otimes y)_{ij} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j y_i y_j = \sum_{i=1}^n x_i y_i \sum_{j=1}^n x_j y_j = (x \cdot y)(x \cdot y).$$

Next lecture, we'll take the $(2p-1)$-element subsets and associate to them a vector in $\mathbb{R}^{n^2}$, and that will allow us to contradict the medium-size intersection result if we have a set with too many elements.

7 February 24, 2022

Last lecture, we started constructing a counterexample to Borsuk's conjecture. As a reminder, the conjecture claims that a set $X \subseteq \mathbb{R}^d$ of finite diameter can be split into $(d+1)$ pieces each with smaller diameter than X (motivated by the arguments for breaking up a d-dimensional ball or a simplex), and what Kahn and Kalai showed is that for $n = 4p$ (p prime), we can construct a finite set of points in $\mathbb{R}^{n^2}$, such that every partition of X into $m < 1.1^n$ pieces, one of the pieces must have the same diameter as the original set X. (So as long as $1.1^n > n^2 + 1$, the conjecture is false, and in fact asymptotically the conjecture is extremely incorrect.)

To start the proof, we mentioned last time that if $x, y \in \mathbb{R}^n$ are two vectors, we can form a vector $x \otimes y \in \mathbb{R}^{n^2}$ whose entries are $x_i y_j$ for $1 \leq i, j \leq n$ (we're not thinking of this as a tensor product, though the entry-wise multiplication is motivated that way). In particular, we noticed that $(x \otimes x) \cdot (y \otimes y) = (x \cdot y)^2$, and we'll now use that to find our finite set of points disproving Borsuk's conjecture.

Proof. Our construction works as follows: remembering our result about medium-size-intersections of sets (specifically $(2p-1)$-element subsets of $\{1, \cdots, 4p\}$), we'll let $\mathcal{A}$ be the family of all $(2p-1)$-element subsets of $\{1, \cdots, 4p\} = \{1, \cdots, n\}$. We have $|\mathcal{A}| = \binom{n}{2p-1} = \binom{4p}{2p-1}$, and now we will associate to every subset in $\mathcal{A}$ a point in $X \subseteq \mathbb{R}^{n^2}$. Since the characteristic vector is n-dimensional, we'll want to use the "tensoring" operation we mentioned above. Furthermore, because the characteristic vector has a lot of zeros (which creates lots of zeros after tensoring), we'll make a slight modification to it as well: let $u_A \in \{1, -1\}^n$ be the **signed characteristic vector**, meaning that u_A has 1 in the jth index if $j \in A$ and -1 otherwise. Then for each $A \in \mathcal{A}$, we define

$$p_A = u_A \otimes u_A \in \mathbb{R}^{n^2}$$

(meaning that the (i, j) entry of p_A is 1 if i, j are both in or both not in A, and -1 otherwise), and we'll let X be the set of all of these points:

$$X = \{p_A : A \in \mathcal{A}\}.$$

We claim that $|X| = |\mathcal{A}| = \binom{n}{2p-1}$, and to show that we need to prove that the p_As are all distinct. Indeed, $(p_A)_{1,j} = (u_A)_1(u_A)_j$, so just looking at the $(1, j)$ entries tells us the vector u_A up to a sign (specifically the sign $(u_A)_1$). And furthermore, u_A will have $(2p-1)$ 1s and $(2p+1)$ -1s, so we can then figure out the sign by counting up the number of 1s in the $(1, j)$ entries of p_A. Thus p_A uniquely determines u_A. (This wouldn't have worked if we had $(2p)$-element subsets – A and A^c would give us the same p_A, for example.)

We now wish to prove our partition property, and to do that we must understand the diameter of X, meaning we need to do some work to understand how distances behave in X. Let $A, B \in \mathcal{A}$. Then

$$\begin{aligned} ||p_A - p_B||^2 &= (p_A - p_B)(p_A - p_B) \\ &= p_A \cdot p_A - 2p_A \cdot p_B + p_B \cdot p_B \\ &= (u_A \otimes u_A) \cdot (u_A \otimes u_A) - 2(u_A \otimes u_A)(u_B \otimes u_B) + (u_B \otimes u_B) \cdot (u_B \otimes u_B), \end{aligned}$$

and now we can use our identity $(x \otimes x) \cdot (y \otimes y) = (x \cdot y)^2$ on every term here to get

$$\begin{aligned} &= (u_A \cdot u_A)^2 - 2(u_A \cdot u_B)^2 + (u_B \cdot u_B)^2 \\ &= n^2 - 2(u_A \cdot u_B)^2 + n^2 \end{aligned}$$

(here we must remember that all entries of u_A and u_B are in $\{-1, 1\}$, not $\{0, 1\}$). Finally, $u_A \cdot u_B$ has entries of 1 for elements in either both or neither of A and B, and entries of -1 in the symmetric difference, meaning that

$$\boxed{u_A \cdot u_B} = |A \cap B| + |\{1, \cdots, n\} \setminus (A \cup B)| - |A \setminus B| - |B \setminus A| = n - 2|A \setminus B| - 2|B \setminus A|$$

(we can draw a Venn diagram to check this, for example). And now because $|A \setminus B| = |A| - |A \cap B|$, and $|A| = 2p-1$ (similar for B as well), we arrive at

$$= n - 2((2p-1) - |A \cap B|) - 2((2p-1) - |A \cap B|) = 4p - 4((2p-1) - |A \cap B|) = \boxed{4(|A \cap B| - (p-1))}.$$

Substituting this back into the norm calculation, we find that

$$\boxed{||p_A - p_B||^2 = 2n^2 - 32\left[|A \cap B| - (p-1)\right]^2},$$

and thus the distance between two points is only dependent on the size of $|A \cap B|$. Furthermore, we now know the diameter of X – the largest possible distance occurs when $[|A \cap B| - (p-1)]^2$ is minimized, and this occurs when $|A \cap B| = p - 1$ (corresponding to a squared distance of $2n^2$).

So now suppose we have a partition of $X = \{p_A : A \in \mathcal{A}\}$ into sets $X_1, \cdots, X_m$ for $m < 1.1^n$. Then there is some part X_i such that $|X_i| \ge \frac{|X|}{m} \ge \frac{\binom{n}{2p-1}}{1.1^n}$ (for example, pick the largest part). And now we can use (the contrapositive of) our corollary from last lecture: let $\mathcal{F}$ be the set of As such that $p_A \in X_i$, so that $X_i = \{p_A : A \in \mathcal{F}\}$. Then $\mathcal{F}$ is a family of $(2p-1)$-element subsets of $\{1, \cdots, n\}$, and **because** $\frac{|\mathcal{F}|}{\binom{n}{2p-1}} = \frac{|\mathcal{X}|}{\binom{n}{2p-1}} \ge 1.1^{-n}$, there must be two members of $\mathcal{F}$ that intersect in exactly $(p-1)$ elements. Then the two corresponding points p_A are separated by the maximum distance in X, and thus (because X_i is a subset of X) $\text{diam}(X_i) = \text{diam}(A)$, as desired. □

We've thus proved that not only is Borsuk's conjecture false for $d+1$ subsets, it's also false if we try to break into $1.1^{\sqrt{d}}$ subsets. It makes sense to ask whether there is a version of Borsuk's conjecture that can be salvaged, and the answer is yes:

Proposition 32

Every set $X \subseteq \mathbb{R}^d$ of finite diameter can be partitioned into at most 7^d subsets such that each subsets has strictly smaller diameter than X.

Proof. Let X be a subset of $\mathbb{R}^d$ with diameter t. Find a maximal set of points $q_1, \cdots, q_m \in X$ of pairwise distance at least $\frac{t}{3}$ (meaning that all points in X are of distance at most $\frac{t}{3}$ to one of those points q_i), and letting $X_j = \left\{x \in X : ||x - q_j|| \le \frac{t}{3}\right\}$, we have $X = X_1 \cup \cdots \cup X_m$. This is not a partition yet, but we can just make the sets smaller until they have no overlap – this only potentially decreases the diameter. Then X_j has diameter at most $\frac{2t}{3} < t = \text{diam}(X)$, so it suffices to prove that $m \le 7^d$. Indeed, the balls of radius $\frac{t}{6}$ around the q_is are disjoint because the q_is are separated by a distance greater than $\frac{t}{3}$, and their union is contained within a ball of radius $t + \frac{t}{6} = \frac{7t}{6}$ from some fixed point $x \in X$ (because all q_i are distance at most t from x, and then points in the balls of radius $\frac{t}{6}$ can only be a further $\frac{t}{6}$ away by the triangle inequality). Since the volume of the big ball is 7^d times larger than the volume of the small balls, and the small balls are disjoint and contained in the big ball, this shows $m \le 7^d$ as desired. □

Fact 33

The next natural question is to ask for the best function $f(d)$ that we can put in Borsuk's conjecture and still have it be true – we have a lower bound of $1.1^{\sqrt{d}}$ and an upper bound of 7^d (in fact $5^d + 1$ if we optimize the constants in the proof). The best known upper bound is $\sqrt{1.5 + o(1)}^d$, and the best known lower bound (that is, how many subsets are necessary) is $(1.225\cdots)^{\sqrt{d}}$. So the only additional information we know is some improvements in the bases of the exponentials – there is still a huge gap in what is known.

Remark 34. *Last lecture, we also had some questions about how important it was for p to be prime in the medium-size-intersections problem, as well as whether we needed the forbidden intersection size to be "middle." In fact, all that really mattered in our proof was how many residue classes mod p are forbidden, so we could have also not allowed either intersections of size 3 or $(p+3)$ for our $(2p-1)$-element subsets, for example. It all basically comes down to modifying the proof from last lecture whenever we're working mod p.*

Meanwhile, if p is not a prime, questions become much more difficult – all of the current methods for proofs rely on linear algebra, and thus there are barely any results known.

We'll close this lecture with a quick riddle (looking ahead to next lecture):

Problem 35

Consider the subset $X = \{(1, 2^1), (2, 2^2), (3, 2^3), \cdots, (10^6, 2^{10^6})\} = \{(j, 2^j) : j \in \mathbb{Z}, 1 \le j \le 10^6\} \subseteq \mathbb{R}^2$. What is a nonzero polynomial $P \in \mathbb{R}[x, y]$ that vanishes on all points in X, what is the minimum possible degree of P?

We can get a degree of 1000000 by using $P(x, y) = \prod_{j=1}^{10^6}(x - j)$, and we can get half that degree (500000) by splitting up the points into pairs, drawing a line through each pair, and multiplying those lines together. We can then go one step further, dividing X into 5-tuples of points, find a quadratic polynomial (conic) through them, and multiply those together to get a degree of 400000. The next step is to use a cubic through nine points, and so on – it turns out that we can actually get a polynomial of degree less than 1500 to work. But we'll talk about this next lecture!

8 March 1, 2022

Last lecture, we discussed the question of finding a polynomial of small degree vanishing at particular points. We'll solve that problem today, starting with a useful result:

Lemma 36

Let $\mathbb{F}$ be a field, and let $X \subseteq \mathbb{F}^n$ be a subset (of n-tuples) of size less than $\binom{d+n}{n}$ for some nonnegative integer d. Then there exists a nonzero polynomial $P \subseteq \mathbb{F}[x_1, \cdots, x_n]$ of degree $\deg P \le d$, such that $P(q) = 0$ for all $q \in X$.

As an example, last lecture we looked at the points $(j, 2^j)$ for $j \in \{1, \cdots, 10^6\}$. Then setting $n = 2$ and $d = 1413$, we get $10^6 < \binom{1415}{2}$, so we can find a degree 1413 polynomial $P(x_1, x_2)$ so that $P(j, 2^j) = 0$ for all $j \in \{1, \cdots, 10^6\}$.

We'll prove this lemma essentially just using linear algebra:

Proof. Let V be the vector space of all polynomials $P \in \mathbb{F}[x_1, \cdots, x_n]$ of degree at most d (this space does contain 0 even though we're not ultimately allowed to use it as our polynomial P) – by stars and bars, the dimension of this space is the number of monomials in $x_1, \cdots, x_n$ of degree at most d, which is $\binom{d+n}{d}$. Suppose $X = \{q_1, \cdots, q_M\}$, so that $\boxed{M = |X| < \binom{d+n}{n}}$ by assumption. Indeed, consider the linear map $V \to \mathbb{F}^M$ which maps P to its values in X:

$$P \mapsto (P(q_1), P(q_2), \cdots, P(q_M)).$$

Our goal is to show that the kernel of this map is nontrivial (since we want $P(q_i) = 0$ for all i). But this is true because the dimension of V is larger than the dimension of $\mathbb{F}^M$ (by the boxed statement above), so there is a nonzero polynomial P with the desired property. □

Remark 37. *Another way to think about this proof is that requiring $P(q) = 0$ is a linear constraint in each of the coefficients of the monomials, and if we have $\binom{d+n}{n}$ coefficients and $|X|$ linear relations, there must be a nonzero choice of coefficients satisfying the relations.*

We'll now explain why this lemma is useful beyond the puzzle that we stated last time – it turns out this result is pretty powerful, and we'll now see how it's applying to the finite field Kakeya and Nikodym problems. But first we'll mention some other useful facts:

Fact 38

If $P \in \mathbb{F}[x]$ is a (one-variable) nonzero polynomial over a field $\mathbb{F}$ and $\deg P \le d$, then there are at most d different elements $a \in \mathbb{F}$ such that $P(a) = 0$.

As a reminder, we basically just factor out the polynomial with the different as. And we're stating this fact with $\deg P \le d$ instead of $\deg P = d$ just for convenience, though it's not really changing the statement because we can always just use a smaller d.

Fact 39

Let q be a prime power, and let $P \in \mathbb{F}_q[x_1, \cdots, x_n]$ be a polynomial of degree at most $(q-1)$ **in each** variable x_i. Then if $P(x_1, \cdots, x_n) = 0$ for all $(x_1, \cdots, x_n) \in \mathbb{F}_q^n$, then P must be the zero polynomial.

In other words, if all exponents of all variables in all monomials of P are smaller than q, then we can't have P vanish everywhere unless P is identically zero. This is one of our homework problems – as a note, we need the restriction on degree because a polynomial like $x_1^q - x_1$ vanishes on all n-tuples of $\mathbb{F}_q^n$ by Fermat's little theorem, but it is not the zero polynomial.

We're now ready to discuss more exciting results, starting with the **finite field Nikodym problem**. The finite field Kakeya problem fell first in 2009, with the methods also working for the finite field Nikodym problem, but we'll talk about them in the opposite order because the technical details are clearer that way:

Definition 40

A set $N \subseteq \mathbb{F}_q^n$ is a **Nikodym set** if for all $x \in \mathbb{F}_q^n$, there is an affine line $\ell \in \mathbb{F}_q^n$ such that $x \in \ell$ and $\ell \setminus \{x\} \subseteq N$.

In other words, for every point $x \in \mathbb{F}_q^n$, there is a line (not necessarily through the origin) ℓ through x such that the entire line, except possibly x itself, is in N. So the set N contains a lot of "almost-complete lines."

Fact 41

Over the real numbers, the condition for being a Nikodym set is the same, though to avoid having the set be too large we restrict the space of consideration to $[0,1]^n \subset \mathbb{R}^n$ – it turns out that there is a Nikodym set of measure zero, even though there are so many almost-complete lines in the set. (On the other hand, it's an open problem to show that the Hausdorff dimension of a Nikodym set $N \subseteq [0,1]^n$ must still be n, and there are connections of this problem to harmonic analysis.)

Some examples of Nikodym sets include $\mathbb{F}_q^n$, or $\mathbb{F}_q^n$ without a point, line, or affine hyperplane (choose our lines away from that hyperplane if x is in the hyperplane and along the hyperplane otherwise). But this set is still pretty large – it has size $q^n - q^{n-1}$ – and we're curious how much smaller we can make a Nikodym set. It turns out that we do need to have "very big" Nikodym sets over finite fields:

Theorem 42

Any Nikodym set $N \subseteq \mathbb{F}_q^n$ has size at least $|N| \ge \binom{q+n-2}{n} \ge \frac{1}{2 \cdot n!} q^n$.

In particular, if we fix the dimension n and consider a large q, we must still have a constant fraction of all points in $\mathbb{F}_q^n$ in our Nikodym set. (And the factor of 2 in the denominator really doesn't matter – it comes from the fact that we have a $q - 1 \ge \frac{q}{2}$ in the expansion of the binomial coefficient.) This is in stark contrast to the real case, and it was very surprising when the result first came out!

Proof. Suppose for the sake of contradiction that there were a Nikodym set of size $|N| < \binom{q+n-2}{n}$. By Lemma 36, there is a nonzero polynomial $P \in \mathbb{F}[x_1, \cdots, x_n]$ of degree at most $(q-2)$ which vanishes on all of N.

We claim that this implies $P(x) = 0$ for all $x \in \mathbb{F}_q^n$. Indeed, for an arbitrary $x \in \mathbb{F}_q^n$; by the properties of a Nikodym set, there is some affine line $\ell \subseteq \mathbb{F}_q^n$ with $\ell \setminus \{x\} \subseteq N$. Let v be a nonzero vector in the direction of ℓ so that we can parameterize $\ell = \{x + tv : t \in \mathbb{F}_q\}$. We then find that

$$x + tv \in \ell \setminus \{x\} \subseteq N \quad \forall t \in \mathbb{F}_q \setminus \{0\},$$

so for all $t \in \mathbb{F}_q \setminus \{0\}$, we have $P(x + tv) = 0$. Now because v and x have been fixed, we can think of our relation as a one-variable polynomial equation in t, and because P has degree at most $(q-2)$ and t shows up linearly in the argument, $P(x+tv)$ is a polynomial in t of degree at most $(q-2)$. But now by Fact 38, because P vanishes on $(q-1)$ points of the form $x + tv$ (for $t \in \mathbb{F}_q \setminus \{0\}$), P must indeed be the zero polynomial in t (**note**: this is not the same as saying that P is the zero polynomial, only that for a fixed x and v it's zero in t), meaning that $P(x + 0v) = P(x) = 0$ as well. This proves the claim.

Finally, applying Fact 39, since $P(x) = 0$ for all $x \in \mathbb{F}_q^n$ and P has degree at most $(q-1)$ in each variable x_i (since P overall only has degree at most $(q-2)$), P must be the zero polynomial, a contradiction. Thus we must have $|N| \geq \binom{q+n-2}{n}$ as desired. □

We'll finish by introducing the finite field Kakeya problem, again using a finite field $\mathbb{F}_q$ for prime power q:

Definition 43

A set $K \subseteq \mathbb{F}_q^n$ is a **Kakeya set** if it contains a line in every direction. In other words, for every vector $a \in \mathbb{F}_q^n \setminus \{0\}$, there is some $b \in \mathbb{F}_q^n$, potentially zero, such that $\{b + ta : t \in \mathbb{F}_q\} \subseteq K$.

Just like before, we're curious how small a Kakeya set may be, and we'll again see that the Kakeya set takes up a constant fraction of the total space for fixed n:

Theorem 44 (Dvir, 2009)

Any Kakeya set $K \subseteq \mathbb{F}_q^n$ has size at least $|K| \geq \binom{q+n-1}{n} \geq \frac{1}{n!} q^n$.

Fact 45

In the real setting, a set is a Kakeya set if it contains a unit interval in every direction, and again in $\mathbb{R}^n$ we have a situation where there are Kakeya sets of measure zero, but it's an open problem to show that the Hausdorff dimension of a Kakeya set must be n.

We'll discuss the proof next time (we've been shifting gears to using more **polynomial methods**), but one note is that finite fields aren't necessarily more rigid than the real numbers – there are often situations where the behavior is very similar. But for a more recent counterexample where we see this disparity, we'll discuss the cap set problem later in the course.

9 March 3, 2022

Today, we'll prove the bound for the finite field Kakeya problem that we described last lecture. Recall that a Kakeya set in $\mathbb{F}_q^n$ (for some prime power q) is a set that contains a line of the form $\{b + ta : t \in \mathbb{F}_q\}$ for any "direction" $a \in \mathbb{F}_q^n$, and we are trying to prove that any Kakeya set has size at least $\binom{q+n-1}{n} \geq \frac{1}{n!} q^n$ (which is a constant fraction of the whole set for fixed n). This proof will look pretty similar to the proof of the finite field Nikodym problem, in which we combined a few lemmas about polynomials.

Proof. Suppose for the sake of contradiction that there existed some Kakeya set with $|K| < \binom{q+n-1}{n}$. Then, analogous to last lecture, we can find a nonzero polynomial $P \in \mathbb{F}_q[x_1, \cdots, x_n]$ of degree $d \le q-1$ which vanishes on all of K. For technical reasons, we'll need to exclude the case where $d = 0$, and indeed this is true because nonzero constants do not vanish on K, and K is nonempty because it needs to contain a line. So P is a polynomial of degree between 1 and $(q-1)$, inclusive.

Our next step in the Nikodym proof was to prove that P vanishes everywhere, but that isn't going to work directly here because the condition looks different. Instead, let's work with what we have: we're told that for every $a \in \mathbb{F}_q^n \setminus \{0\}$, there is a vector $b \in \mathbb{F}_q^n$ such that $\{b + ta : t \in \mathbb{F}_q\} \subseteq K$. Thus, $P(b + ta) = 0$ for all $t \in \mathbb{F}_q$, and if we look at this as just a polynomial in t with fixed a, b, we have q roots (all $t \in \mathbb{F}_q$) but only degree **at most** $d \le q-1$ (again, remember that the degree in t can be less than the degree of P itself because of cancellation), so (applying another fact from last time) $P(b + ta)$ must be the zero polynomial in t. (Remember that over finite fields, polynomials like $(x^q - x)$ mean that vanishing everywhere does not tell us that we have the zero polynomial – the degree constraint is indeed necessary.)

We can now look at the t^d coefficient in $P(b+ta)$, which we know must be zero for any a. Since P is itself degree d, we can only use the leading terms, and we'll write

$$P(x_1, \cdots, x_n) = P_d(x_1, \cdots, x_n) + P_{\le d-1}(x_1, \cdots, x_n),$$

where the first term P_d is the homogeneous degree d part of P (and $P_{\le d-1}$ is the rest of the terms of P, which all have degree at most $(d-1)$). Since we assume P actually has degree d, we know that P_d is a nonzero polynomial. Plugging in our argument $b+ta$ into the polynomial P, notice that $P_{\le d-1}$ does not contribute any t^ds (because it only has degree at most $(d-1)$), and furthermore we can't use any bs in the first term for t^d terms because we must use the part with t every time. (To be more clear, we can write $b + ta$ as $(b_1 + ta_1, b_2 + ta_2, \cdots, b_n + ta_n)$ and imagine expanding it out.)

Thus the contribution to the t^d coefficient is just $P_d(ta) = t^d P_d(a)$, and remembering that this is zero, we find that $P_d(a) = 0$ **for all nonzero** $a \in \mathbb{F}_q^n$ (remembering that a has to be a line direction). And $P_d(0) = 0$ as well, because P_d is homogeneous of degree $d \ge 1$. Thus P_d vanishes everywhere and has degree $d \le q-1$, so it definitely has degree at most $q-1$ in each variable and (by our homework lemma) P_d is the zero polynomial. This is a contradiction, proving that our initial assumption of $|K| < \binom{q+n-1}{n}$ was incorrect. □

Fact 46

The best known bound for this problem (Bukh–Chao 2021) is that $|K| \ge \frac{1}{(2-1/q)^{n-1}} q^n$ (so basically we can improve our $\frac{1}{n!}$ factor to $\frac{1}{2^n}$). On the other hand, there is a known construction (published in 2008, basically right after Dvir's proof) of a Kakeya set $K \subseteq \mathbb{F}_q^n$ of size $\frac{1}{2^{n-1}} q^n + O_n(q^{n-1})$ (O_n meaning asymptotics for n fixed, q large). So we basically have matching lower and upper bounds here!

Fact 47

On the other hand, results for the finite field Nikodym problem are more sparse, and even the smallest known constructions have size $(1 - o(1))q^n$ (such as our example from last time deleting a hyperplane from $\mathbb{F}_q^n$). And this has been conjectured to be optimal.

We'll now turn to the **joints problem**, another geometrically-motivated combinatorics problem with ideas motivated by the Kakeya and Nikodym problems, but living in $\mathbb{R}^3$ instead of $\mathbb{F}_q^n$. We'll start with some easier riddles to motivate the statement:

Problem 48

If we have a set of L distinct lines in the plane, what is the maximum possible number of points that lie on at least two of the lines?

(The answer is $\binom{L}{2}$, because any two lines can only intersect in at most one point, and we can achieve this by choosing lines generically.)

Problem 49

If we have a set of L distinct lines in the plane, what is the maximum possible number of points that lie on at least **three** of the lines?

We can get $O(L)$ points by picking $\frac{L-1}{2}$ points on one of the lines and having pairs of lines pass through each of those points, and we can improve that to $O(L^2)$ by drawing a grid of horizontal, vertical, and down-right diagonal lines (or equivalently drawing an equilateral triangular grid). And we have an upper bound of $\frac{1}{3}\binom{L}{2}$ because each of the points must be counted in at least three pairs of points from the previous problem.

Problem 50

If we have a set of L distinct lines **in** $\mathbb{R}^3$, what is the maximum number of points that lie on at least three of the lines?

This problem isn't more interesting than the previous one – we can still take the same example as above by just restricting to a plane, and $O(L^2)$ is still possible and still required. But to make the problem more interesting, we can try to reduce this degeneracy and not allow the lines to be coplanar. That finally leads us to the problem:

Definition 51

Let $\mathcal{L}$ be a set of lines in $\mathbb{R}^3$. A **joint** of $\mathcal{L}$ is a point $p \in \mathbb{R}^3$ which lies on three non-coplanar lines of $\mathcal{L}$.

It's okay if there are three lines through p that are coplanar, as long as we can find three that are not. And it's important that a joint is a **point**, not a triple of lines (otherwise we could just have all lines through the origin and the problem would not be interesting). And this finally leads us to the joints problem:

Problem 52

Let $L \in \mathbb{N}$. What is the maximum number of joints of a set $\mathcal{L}$ of L lines in $\mathbb{R}^3$?

This problem was first posed in the 1990s with an upper bound of $L^{7/4}$, and it wasn't until 2010 that Guth (Professor Larry Guth in our department) and Katz resolved it. We'll start with some constructions: if we take a $k \times k \times k$ grid in $\{1, \cdots, k\}^3$, we need to use $L = 3k^2$ lines and we get k^3 joints, so this allows us to achieve $O(L^{3/2})$ (with a constant of $\frac{1}{\sqrt{27}}$.

To get a slightly better construction (in fact best known), we can consider k planes in $\mathbb{R}^3$ in general position (meaning that we don't get extra intersections, for example) and let the lines in $\mathcal{L}$ be the $L = \binom{k}{2}$ intersection lines between any two planes. We then get $\binom{k}{3}$ joints (from the intersection of any three planes), meaning that we again get $O(L^{3/2})$ but now with an improved constant of $\frac{\sqrt{2}}{3}$. And it turns out this is the best possible up to constant factors:

Theorem 53 (Guth–Katz 2010)
If $\mathcal{L}$ is a set of L lines in $\mathbb{R}^3$, then $\mathcal{L}$ has at most $3L^{3/2}$ joints.

Fact 54
The constant 3 in the theorem above can be optimized, as we'll see in the proof, but it turns out the true constant is actually $\frac{\sqrt{2}}{3}$ as in our second construction above (as proved by Yu and Zhao in 2019).

We won't have time to do the whole proof today, but we'll mention the technical details today and do the main argument next lecture:

Lemma 55
Let $\mathcal{L}$ be a set of lines in $\mathbb{R}^3$ (of any size), and suppose that $\mathcal{L}$ has J joints. Then some line in $\mathcal{L}$ contains at most $2J^{1/3}$ joints.

We'll prove this lemma next time, but for now we can see why it implies Theorem 53:

Proof of Theorem 53 assuming Lemma 55. Let $f(L)$ be the maximum possible number of joints of L lines; our goal is to show that $f(L) \le 3L^{3/2}$. (Notice that $f(0) = 0$ and that $f(0) \le f(1) \le f(2) \le \cdots$.) **We claim that** $f(L) \le f(L-1) + 2f(L)^{1/3}$; indeed, if $\mathcal{L}$ is a set of lines in $\mathbb{R}^3$ of maximally many ($J = f(L)$) joints, then one of the lines contains at most $2J^{1/3} = 2f(L)^{1/3}$ joints. Deleting this gives us at most $f(L-1)$ joints left, and we could have only removed at most $2f(L)^{1/3}$ of them, proving the claim.

From here, repeatedly applying the claim gives us

$$f(L) \le f(L-1) + 2f(L)^{1/3} \le f(L-1) \le f(L-2) + 2f(L-1)^{1/3} + 2f(L)^{1/3} + \cdots,$$

finally arriving at (using $f(0) = 0$)

$$\boxed{f(L)} \le 2f(1)^{1/3} + 2f(2)^{1/3} + \cdots + 2f(L)^{1/3} \le \boxed{L \cdot 2f(L)^{1/3}},$$

so that $f(L)^{2/3} \le 2L$ and thus $f(L) \le 2^{3/2}L^{3/2} \le 3L^{3/2}$. □

10 March 8, 2022

Last lecture, we set up the **joints problem**, which aims to find the maximum number of joints (points that lie on three non-coplanar lines) formed by a set of L lines. We're aiming to prove Guth and Katz's 2010 result, which proves a bound of $3L^{3/2}$ joints (a better constant can be achieved but we won't discuss it), and we mentioned that the proof follows from the **main lemma** stated last time, namely that if a set of lines $\mathcal{L}$ (no constraints on the number of them) has exactly J joints, then one of the lines has at most $2J^{1/3}$ joints.

Proof of main lemma. (If $J = 0$ this is clear.) Assume for the sake of contradiction that all lines contain more than $2J^{1/3}$ joints. Much like in the proof of the finite field Kakeya problem, we will try to find a polynomial that vanishes on all joints. Since we don't know what the optimal degree is, we choose a nonzero polynomial $P \in \mathbb{R}[x, y, z]$ of **minimum possible degree**, such that $P(q) = 0$ for all joints q.

We claim that we can make $\deg P \le 2J^{1/3}$. Indeed, recall our lemma that given a subset $X \subseteq \mathbb{F}^n$ of size less than $\binom{d+n}{n}$, there is a nonzero polynomial P of degree at most d such that $P(q) = 0$ for all $q \in X$. So because $n = 3$ here, if we plug in $d = 2J^{1/3}$ into the lemma, then we find that $\binom{d+n}{n} = \binom{\lfloor 2J^{1/3} \rfloor + 3}{3} > \frac{(2J^{1/3})^3}{6} > J$. So because there are J joints, we do indeed have a polynomial of degree at most d with the vanishing that we want.

But now P vanishes on more points on each line than $\deg P$, so on each line P must vanish completely (since restricted to that line, by parameterization, P is a one-variable polynomial with more roots than its degree and thus must be the zero polynomial). And now **here's a new idea that hasn't come up in proofs before**: since P vanishes on lines of the form $\{a + tb : t \in \mathbb{R}\}$, the directional derivative of P in the direction of b vanishes on the entire line as well (since P is constant on that line). Thus,

$$\nabla P(a) \cdot b = 0 \quad \text{if } \{a + tb : t \in \mathbb{R}\} \in \mathcal{L}.$$

But now we can use the fact that whenever a is a joint, we have three different bs which satisfy this equality (corresponding to the three lines through a) Furthermore, these three bs can be chosen to span all of $\mathbb{R}^3$ (by definition). So $\nabla P(a)$ can be dotted with three different non-coplanar vectors, resulting in 0 in all cases, and thus $\nabla P(a) = 0$ **for any joint** a of $\mathcal{L}$.

However, ∇P is a vector of three entries $\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y}, \frac{\partial P}{\partial z}$, and we've just found that each of these polynomials vanishes on all of the joints. But those polynomials are of lower degree than P, and we initially chose P to be of minimum degree. Thus, the only way this could be possible is if each of those polynomials is the zero polynomial, but that would require P to only be a function of y and z, and also only a function of x and z, and also only a function of x and y, which only happens if P is constant and that's a contradiction when $J > 0$. (We do need all three polynomials here to get our contradiction!) Thus there must be some line which contains at most $2J^{1/3}$ points. □

This concludes the part of the class discussing polynomial methods relying on the number of roots a polynomial can have based on its degree. We'll now spend a bit of time discussing **spectral methods** – we'll still be using linear algebra, but instead of just making arguments about linear independence and so on, we'll be considering properties of matrices like eigenvalues.

Definition 56

Let G be a graph with vertices $\{1, \cdots, n\}$. The **adjacency matrix** of G is the $n \times n$ matrix $A = (a_{ij})$, where a_{ij} is 1 if i and j are adjacent and 0 otherwise.

While this matrix has a natural definition which doesn't seem to give us much illuminating information about G, it's surprising how much we can actually gain from looking at the eigenvalues of this matrix. But first, we'll make some observations: A has all entries 0 or 1 and with zeros on the diagonal, so $\mathrm{tr}(A) = 0$ (and thus the sum of the eigenvalues is also zero). And because we have a real symmetric matrix, the spectral theorem tells us that A has all real eigenvalues and that we can find an orthogonal basis of eigenvectors spanning $\mathbb{R}^n$ (we do not even need to consider generalized eigenspaces or Jordan normal form).

Lemma 57

Let G be a d-regular graph (so $\deg(v) = d$ for all vertices v). Then d is an eigenvalue of the adjacency matrix A of G, and all other eigenvalues λ of A satisfy $|\lambda| \le d$.

Proof. Since G is d-regular, the adjacency matrix has exactly d 1s in each row. Thus, the all-ones vector will be an eigenvector of A with eigenvalue d (since dotting the all-ones vector with any row gives us d). For the other eigenvalues, suppose λ is an eigenvalue of A, and let $v = (v_1, \cdots, v_n)^T$ (a column vector) be a nonzero eigenvector of this eigenvalue λ. Let j be an entry such that $|v_j|$ is maximal, so that $|v_i| \le |v_j|$ for all i and $|v_j| > 0$ (because v is nonzero). If we then look at the jth coordinate of $Av = \lambda v$, then we get

$$\lambda v_j = \sum_{i \sim j} v_i,$$

where $i \sim j$ means that i is a neighbor of j in the graph G. But now taking absolute values and using the triangle inequality, we have

$$|\lambda||v_j| = \left|\sum_{i \sim j} v_i\right| \le \sum_{i \sim j} |v_i| \le d|v_j|$$

because j is adjacent to exactly d other vertices. Thus because $|v_j| > 0$, dividing by it yields $|\lambda| < d$, as desired. □

(In particular, the eigenvalue $-d$ shows up if and only if the graph is bipartite, which we'll show on homework.) And we'll now see with this next result that we can do matrix calculations with the adjacency matrix, explaining why this is a good way to encode the information about the edges of G:

Lemma 58

Let G be a graph with adjacency matrix A. Then the (i, j) entry of A^m is the number of walks of length m from i to j (paths that are allowed to repeat vertices and edges).

(For example, a walk 1 → 17 → 5 → 1 → 3 is of length 4, so the $(1, 3)$ entry of A^4 will get a contribution from this walk if all adjacent vertices in that walk are edges of G.) This is basically an exercise in thinking about how matrix multiplication works with adjacency matrices, and it'll be on our homework as well.

We'll now look at a graph theory result, the **friendship theorem**, which was originally proved without spectral methods but where the proof using adjacency matrices turns out to be simpler.

Theorem 59 (Friendship theorem)

Let G be a graph such that any two distinct vertices have exactly one common neighbor. Then the only valid graphs G are sets of triangles all joined together at a common vertex (we can check that any two vertices here indeed share a neighbor), which we call "windmill graphs."

For example, we can imagine that the graph encodes the friendships between a group of people, and any two people are exactly one common friend. Then in particular, this result tells us that there is one person who is friends with everyone else, and then we have a perfect matching of the remaining people. (This result was originally proved by Erdös, Rényi, and Sós in 1966.)

Start of proof. Suppose there is a vertex connected to all other vertices. Then applying the theorem condition to the central vertex and any other vertex, we find that that other vertex must have degree 1 among the other remaining vertices. Repeating this logic, the graph induced by the remaining vertices has every degree 1, so it must be a perfect matching and that gives us the windmill graph situation. So the theorem is satisfied in this case.

So now we can assume that there is a vertex not connected to all other vertices. We wish to show that the graph must be regular in this case, and we'll start by getting halfway there. **We claim that** for any vertices x, y that are not adjacent in G, we have $\deg(x) = \deg(y)$. Indeed, let the neighbors of x be $v_1, \cdots, v_m$. Applying the theorem condition to v_i and y, we get a common neighbor v_i' of v_i and y which is not x. Furthermore, v_1' through v_m' are all distinct vertices (if v_i' and v_j' were both the same, then v_i and v_j would share both x and $v_i' = v_j'$ as neighbors, which is not allowed). Thus y has at least as many neighbors as x. Flipping the logic shows the result the other way around as well, so $\deg(x) = \deg(y)$. □

We'll continue this proof next time!

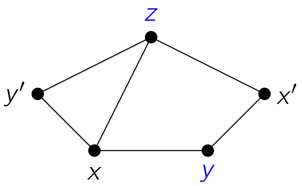
11 March 10, 2022

Last time, we started discussing the friendship theorem, which we're trying to prove using spectral graph theory. Recall that the setup is as follows: suppose G is a graph such that any two vertices have exactly one common neighbor. We wish to prove that G must be a "windmill graph," where one vertex is connected to all other vertices and the remaining vertices induce a perfect matching. What we've already shown last time is that if G has a vertex that is adjacent to all other vertices, then G is indeed a windmill graph. Furthermore, we've considered the other case (where any vertex of G is not adjacent to all other vertices), and we've shown that for any vertices x, y that are not adjacent, then $\deg(x) = \deg(y)$ (by showing that we can associate any neighbor of x with a different neighbor of y, and vice versa). This is nice because spectral graph theory is particularly nice for regular graphs, and now we'll continue the proof:

Continuation of the proof of the friendship theorem. We now know that if no vertex of G is adjacent to all other vertices, then G has many equal degrees among its vertices. We'll now take this a step further and **claim that G is in fact regular**. Indeed, suppose we have two vertices x, y that are adjacent in our graph. If we could show that there is some vertex z that is not adjacent to either x or y, then the previous claim would give us $\deg(x) = \deg(z)$ and $\deg(y) = \deg(z)$, so that $\deg(x) = \deg(y)$. So the only remaining case is the one where every vertex is adjacent to either x or y.

We're now in the case where no vertex of G is adjacent to all other vertices, so in particular, there is some vertex x' which is not adjacent to x and some vertex y' which is not adjacent to y. Notice that $x' \neq y$ and $y' \neq x$ because we chose x and y to be adjacent, and in fact x' is adjacent to y (since it's not adjacent to x but needs to be adjacent to either x or y) and y' is adjacent to x. Furthermore, x' and y' are not the same vertex because they have different relations to x and y.

Thus, we have a picture of y', x, y, x' connected in a path in that order in G, and using the friendship theorem condition, x' and y' have a common neighbor z which is none of x, y, x', y'. This means we have formed a pentagon with the five vertices x, y, x', z, y' in that order, but because z is not x or y it must be connected to one of those vertices by assumption. Without loss of generality, assume z is adjacent to x. And this is a contradiction, because now z and y have two common neighbors (x and x') and they should only have one. Thus in this case we must have a vertex z adjacent to neither x nor y and we indeed have $\deg(x) = \deg(z) = \deg(y)$.



Bringing things back now, we now know that if G doesn't have a "central vertex" connected to everything else, then G is regular because all degrees are equal (call this common degree d). If $d = 0$ the theorem statement is true (we must have the one-vertex graph), and if $d = 1$ we have a perfect matching and G cannot exist. So $d \geq 2$, and from here we will actually finish the proof pretty quickly using spectral graph theory. Suppose we have n vertices, and let A be the $n \times n$ adjacency matrix of G. Then A^2's (i, j) entry is the number of walks from i to j in exactly two steps, which is d for each diagonal entry (i, i) (walk to a neighbor of i and back) and 1 for each off-diagonal entry (i, j) (use the only common neighbor of i and j). (Here is where it's important that we're talking about walks, not paths!)

Thus, A^2 is of the form

$$A^2 = \begin{bmatrix} d & 1 & \cdots & 1 \\ 1 & d & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & d \end{bmatrix} = (d-1) \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

(notice that it's nice that even though we barely know any of the entries of A, we know all of A^2 explicitly), and now we can compute the eigenvalues of A^2: the entire $(n-1)$-dimensional subspace $x_1 + \cdots + x_n = 0$ is in the kernel of the all-ones matrix, so A^2 has an eigenvalue of $(d-1)$ with multiplicity $(n-1)$. Then the final eigenvalue is $n+d-1$ (coming from the all-ones vector; we could also have found this by computing the trace of A^2).

We thus know the eigenvalues of A up to a sign: they're the $(\pm)$ square roots of the eigenvalues of A^2. But remember that for any d-regular graph (which A is), one of the eigenvalues of A must be d. And $\sqrt{d-1} < d$ for $d \geq 2$, so the only possibility is $\sqrt{n+d-1} = d$. It turns out the specific value of n doesn't matter: instead, we find that the eigenvalues of A are $d = \sqrt{n+d-1}$ and $(n-1)$ eigenvalues that are each one of $\pm\sqrt{d-1}$. Furthermore, the trace of A is zero, so we have

$$d + \alpha\sqrt{d-1} - \beta\sqrt{d-1} = 0$$

for some $\alpha + \beta = (n-1)$. But letting $\beta - \alpha = k$, we find that

$$d = k\sqrt{d-1} \implies k^2(d-1) = d^2,$$

which implies that $(d-1)|d^2 \implies (d-1)|d^2 - (d-1)(d+1) = 1$, which can only occur if $d = 2$ (since we assumed $d \geq 2$). So we have a 2-regular graph of 3 vertices, and that is in fact a situation where G has a vertex adjacent to all other vertices (this is the single-triangle windmill graph). (Alternatively, if we didn't know the value of n, we could note that a 2-regular graph is a collection of cycles which needs to be connected.) This concludes the proof – even in this case, the only valid G is a windmill graph. □

We'll now move on to a powerful result, which concerns **eigenvalue expansion properties of graphs** (in other words, instead of employing spectral graph theory to prove a result, we base our result off of spectral graph theory from the start). Intuitively, a graph is an **expander** if it is very well-connected, meaning that between any subset of the vertices and its complement, there are a lot of edges (so we don't have a situation where there's one half of the graph connected to the rest by only a single edge). We won't go into the details of those definitions precisely, but we will talk about how we can understand expansion properties from eigenvalues.

Here's the setup: let G be a d-regular graph on n vertices, let A be its adjacency matrix, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues (with multiplicity) of A. The spectral theorem gives us an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors $v_1, \cdots, v_n$ (meaning that $Av_i = \lambda_i$ for all i). We know that $\lambda_1 = d$ and $\lambda_2 \geq \cdots \geq \lambda_n \geq -d$: we will use the notation that v_1 is always the normalized all-ones vector (with $\frac{1}{\sqrt{n}}$ in each entry), even when d is an eigenvalue of higher multiplicity.

Definition 60

For a d-regular graph G and subsets of vertices $B, C \subseteq V(G)$ (not necessarily disjoint), define the number of edges between B and C to be

$$e(B, C) = |\{(u, w) \in B \times C : u, w \text{ adjacent}\}| .$$

(In particular, edges contained within both B and C are counted twice, and if $C = B^c$ then we do have the ordinary number of edges between B and its complement.) We can in fact compute this quantity using the adjacency matrix, because

$$e(B, C) = 1_B^T A 1_C$$

for the characteristic vectors $1_B, 1_C$ of B and C (we get a 1 contribution whenever there's a 1 in some entry of B, a 1 in some entry of C, and then a 1 in the corresponding spot in A that connects those vertices in B and C).

Lemma 61

Let G be a d-regular graph on n vertices. Then for any subset $B \subset V(G)$, we have

$$e(B, V(G) \setminus B) \ge (d - \lambda_2)\frac{|B|(n - |B|)}{n}.$$

In other words, the number of edges required between B and the rest of the graph can be characterized by the **spectral gap** between the first and second eigenvalues of the adjacency matrix of G (this inequality is useful whenever $\lambda_2 < d$ and we don't have a multiple eigenvalue at d). To prove this, we'll make use of the following elementary fact:

Fact 62

From linear algebra, we know that because $v_1, \cdots, v_n$ form an orthonormal basis, any $w \in \mathbb{R}^n$ can be written as $w = a_1 v_1 + \cdots + a_n v_n$ for $a_i = w \cdot v_i$, so that $||w||^2 = a_1^2 + \cdots + a_n^2$.

Proof of lemma. Let $\overline{B}$ denote the complement $V(G) \setminus B$ of B for convenience, and let 1_B and $1_{\overline{B}}$ be the characteristic vectors of B and $\overline{B}$, respectively, Notice that $1_B + 1_{\overline{B}}$ is the all-ones vector. Then

$$e(B, \overline{B}) = 1_B^T A 1_{\overline{B}} = 1_B (A 1_{\overline{B}}).$$

Using the fact above, we can write our characteristic vector in terms of the orthonormal basis of eigenvectors

$$\boxed{1_B = b_1 v_1 + \cdots + b_n v_n},$$

where in particular $b_1 = 1_B \cdot v_1 = \frac{|B|}{\sqrt{n}}$ (because v_1 is $\frac{1}{\sqrt{n}}$ times the all-ones vector). Additionally, because $1_{\overline{B}}$ is the all-ones vector minus 1_B, which is $\sqrt{n} v_1 - 1_B$, we have

$$1_{\overline{B}} = \left(\sqrt{n} - \frac{|B|}{\sqrt{n}}\right) v_1 - b_2 v_2 - \cdots - b_n v_n.$$

Now the action of A on $1_{\overline{B}}$ is simple to write down, because each term is an eigenvector of A:

$$\boxed{A 1_{\overline{B}} = \lambda_1 \frac{n - |B|}{\sqrt{n}} v_1 - \lambda_2 b_2 v_2 - \cdots - \lambda_n b_n v_n}.$$

Dotting this with 1_B and using orthonormality of the v_is now gives us

$$e(B,\overline{B}) = 1_B^T A 1_{\overline{B}} = \lambda_1 \frac{|B|}{\sqrt{n}} \cdot \frac{(n-|B|)}{\sqrt{n}} - \lambda_2 b_2^2 - \lambda_3 b_3^2 - \cdots - \lambda_n b_n^2.$$

But because $\lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n$ and $\lambda_1 = d$, we can upper bound this by

$$\geq d\frac{|B|(n-|B|)}{n} - \lambda_2(b_2^2 + \cdots + b_n^2),$$

and now using that $b_2^2 + \cdots + b_n^2 = ||1_B||^2 - b_1^2 = |B| - \frac{|B|^2}{n} = \frac{|B|(n-|B|)}{n}$, we can simplify this as

$$e(B,\overline{B}) = \frac{|B|(n-|B|)}{n} - \lambda_2\frac{|B|(n-|B|)}{n} = (d-\lambda_2)\frac{|B|(n-|B|)}{n},$$

as desired. □

12 March 15, 2022

Last lecture, we discussed expansion properties that we can deduce from the eigenvalues of the adjacency matrix of a graph. Specifically, we know that the largest eigenvalue of a d-regular graph's adjacency matrix is d (corresponding to a normalized all-ones vector), and then we can count the number of edges between a subset of the vertices and the rest of them in terms of the spectral gap $d - \lambda_2$:

$$e(B, V(G) \setminus B) \geq (d-\lambda_2)\frac{|B|(n-|B|)}{n}.$$

We may ask whether there's a corresponding upper bound in terms of the spectral gap, and we may recall that we bounded a bunch of eigenvalues by $\lambda_2, \cdots, \lambda_n \leq \lambda_2$ (this was the only inequality we had in the proof). So we can do the same thing but replacing λ_2 with λ_n to get an upper bound, but that's not a very good bound because λ_n is typically far and we already have the trivial bound $e(B, V(G) \setminus B) \leq d \cdot \min(|B|, n-|B|)$.

But now we'll generalize and ask about $e(B,C)$ more generally (recall that this is the number of ordered pairs $(u,w) \in B \times C$ with u and w adjacent), and an upper bound might be more interesting in this case. If we imagine a d-regular bipartite graph and we set B and C to be subsets both within the same part, then $e(B,C) = 0$; furthermore, the graph is bipartite if and only if $\lambda_n = -d$ (we will show this on our homework). In other words, **often we want to control both the distance between λ_2 and d and the dsitance between λ_n and $-d$** if we want to get upper bounds.

Lemma 63

Let G be a d-regular graph on n vertices with adjacency matrix eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Suppose there is some $\lambda \geq 0$ such that $|\lambda_i| \leq \lambda$ for all $2 \leq i \leq n$ (equivalently, $\lambda_2 \leq \lambda$ and $\lambda_n \geq -\lambda$). Then for any subsets $B, C \subseteq V(G)$, we have

$$\left| e(B,C) - \frac{d}{n}|B||C| \right| \leq \lambda\sqrt{|B||C|}.$$

The idea here is that the density of the graph is around $\frac{d}{n}$ (if we pick two vertices independently at random, the probability that they are adjacent is $\frac{d}{n}$). So we should expect that among the pairs $(u,w) \in B \times C$, about $\frac{d}{n}|B||C|$ of them exist on average. And this result tells us that we are indeed close to this mean, controlled by λ. (But this is only really useful when λ is significantly smaller than d – for example, in the case where we have the d-regular bipartite graph and $e(B,C) = 0$, the inequality reduces to $|B| \cdot |C| \leq n^2$, which is useless.)

Proof. As usual, let $1_B, 1_C \in \mathbb{R}^n$ be the characteristic (indicator) vectors of B and C, respectively. We again have $e(B, C) = 1_B^T A 1_C$, and again we'll express the indicator random variables in terms of the orthonormal eigenvector basis:

$$1_B = b_1 v_1 + \cdots + b_n v_n, \quad b_1 = 1_B \cdot v_1 = \frac{|B|}{\sqrt{n}}$$

(this is repeating the same calculation as last lecture – remember that $b_1^2 + \cdots + b_n^2 = ||1_B||^2 = |B|$) and similarly

$$1_C = c_1 v_1 + \cdots + c_n v_n, \quad c_1 = \frac{|C|}{\sqrt{n}}$$

and $c_1^2 + \cdots + c_n^2 = |C|$. Then plugging into the formula for $e(B, C)$,

$$e(B, C) = 1_B^T A 1_C = 1_B \cdot (A 1_C) = (b_1 v_1 + \cdots + b_n v_n) \cdot (\lambda_1 c_1 v_1 + \cdots + \lambda_n c_n v_n)$$

and now by orthonormality this simplifies to

$$= \lambda_1 b_1 c_1 + \lambda_2 b_2 c_2 + \cdots + \lambda_n b_n c_n.$$

The first term here is $d \cdot \frac{|B|}{\sqrt{n}} \cdot \frac{|C|}{\sqrt{n}} = \frac{d}{n}|B||C|$, which is a good sign, and now we want to bound the remaining terms. Subtracting off the first term from the rest, we now have

$$e(B, C) - \frac{d}{n}|B||C| = \lambda_2 b_2 c_2 + \cdots + \lambda_n b_n c_n,$$

so that

$$\left| e(B, C) - \frac{d}{n}|B||C| \right| \le \lambda(|b_2||c_2| + \cdots + |b_n||c_n|)$$

(this is where we use the assumption that $|\lambda_i| \le \lambda$ for all i). And now by Cauchy-Schwarz, this simplifies to

$$\le \lambda \sqrt{b_2^2 + \cdots + b_n^2} \sqrt{c_2^2 + \cdots + c_n^2} \le \lambda \sqrt{|B|} \sqrt{|C|},$$

which is precisely the result we want. □

(Notice that in the last inequality we've ignored the effect of the $b_1^2 = \frac{|B|^2}{n}$ and $c_1^2 = \frac{|C|^2}{n}$ in the square roots – if we kept those in our consideration, we'd end up with a very similar to result to the one we proved last lecture.) So the moral of the story here is that having good information about eigenvalues can tell us interesting information about edges in a graph.

We're now going to see how these kinds of methods can be applied to a recently proved and celebrated result, **the sensitivity conjecture**. But we'll have to do some more preparation for that first:

Theorem 64 (Min-max principle)

Let A be a real symmetric $n \times n$ matrix (for example, an adjacency matrix), and let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ be its eigenvalues. Then for any $1 \le j \le n$, we can write the jth eigenvalue in two different ways:

$$\lambda_j = \max_{\substack{|U| \subseteq \mathbb{R}^n \\ \dim U = j}} \left(\min_{\substack{x \in U \\ ||x|| = 1}} (Ax) \cdot x \right) = \min_{\substack{U \subseteq \mathbb{R}^n \\ \dim U = n - j + 1}} \left(\max_{\substack{x \in U \\ ||X|| = 1}} (Ax) \cdot x \right)$$

In other words, we either find the j-dimensional subspace for which we can always guarantee the best lower bound for $(Ax) \cdot x$, or we find the $(n - j + 1)$-dimensional subspace for which we can find the best universal upper bound for $(Ax) \cdot x$.

Proof. Let $v_1, \cdots, v_n$ be an orthonormal basis of eigenvectors of A with corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$. Any vector x can be represented as $c_1 v_1 + \cdots + c_n v_n$, and

$$(Ax) \cdot x = (\lambda_1 c_1 v_1 + \cdots + \lambda_n c_n v_n)(c_1 v_1 + \cdots + c_n v_n) = \lambda_1 c_1^2 + \cdots + \lambda_n c_n^2.$$

Since x is always of norm 1 in the statement of the min-max principle, we must have $c_1^2 + \cdots + c_n^2 = 1$. So we can think of each choice of x as a weighing factor between the λ_is, and $(Ax) \cdot x$ is always just a weighted average of the λ_is (in particular it is always between λ_1 and λ_n).

We'll show that both the max-min and min-max expressions are equal to λ_j by showing both directions of inequalities:

- For the max-min term, to prove the $\leq$ direction, we can consider the subspace $U = \text{span}(v_1, \cdots, v_j)$. Then we can only make use of the coefficients c_1 through c_j, so $(Ax) \cdot x$ (for unit vector $x \in U$) is always a weighted average between $\lambda_1, \cdots, \lambda_j$, and thus the minimum value is λ_j. This means we've found a subspace U where the parenthetical term on the right-hand side is λ_j, so the maximum is indeed at least λ_j.

- Similarly, we can prove the $\geq$ direction for the min-max term by picking $U = \text{span}(v_j, \cdots, v_n)$.

- It might seem like the other direction is more difficult because we have to look at general subspaces U, but it turns out to be not too bad. Suppose for the sake of contradiction that the min-max term were strictly smaller than λ_j, so that there is some subspace $U^* \subseteq \mathbb{R}^n$ of dimension $n - j + 1$ with $\max_{\substack{x \in U^* \\ ||x||=1}} (Ax) \cdot x < \lambda_j$ – in other words, $\boxed{(Ax) \cdot x < \lambda_j}$ for all unit-norm $x \in U^*$. Then if we consider $U = \text{span}(v_1, \cdots, v_j)$, a j-dimensional space, we know that U and U^* have a common vector because the sum of their dimensions is $n + 1$. But we proved (in the first bullet point) that $\boxed{(Ax) \cdot x \geq \lambda_j}$ for all unit-norm $x \in U$, so we arrive at a contradiction and thus we must have equality.

- A similar argument works for proving the equality in the max-min term by intersecting a problem subspace with $\text{span}(v_j, \cdots, v_n)$.

□

We'll now mention another result that we'll make use of towards the sensitivity conjecture:

Theorem 65 (Cauchy interlace theorem)

Let A be a real symmetric $n \times n$ matrix, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues. Let B be an $(n-1) \times (n-1)$ matrix obtained by deleting the kth row and kth column of A (which is notably still symmetric); call its eigenvalues $\mu_1 \geq \cdots \geq \mu_{n-1}$. Then the eigenvalues interlace:

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \cdots \geq \mu_{n-1} \geq \lambda_n.$$

(This may remind us of Rolle's theorem from calculus, which tells us that the roots of a polynomial f and its derivative f' interlace. But the two results aren't really related.) Our proof will make use of the min-max principle:

Proof. Without loss of generality, we'll assume that we delete the nth row and nth column. (We can do this because the eigenvalues of A do not change if we apply the same permutation to the rows and columns of A – the eigenvectors get similarly permuted but the eigenvalues stay the same.) If we delete the nth row and the nth column, it's now natural to embed $\mathbb{R}^{n-1}$ into the first $(n-1)$ coordinates of $\mathbb{R}^n$: for any $y \in \mathbb{R}^{n-1}$, consider $x = (y, 0)$. We know that $||x|| = ||y||$ and also that

$$(By) \cdot y = (Ax) \cdot x,$$

because B is the top left $(n-1)\times(n-1)$ submatrix of A and the additional entries of A don't do anything if the nth entry of x is zero. So now by Theorem 64 on the matrix B and then embedding $\mathbb{R}^{n-1}$ into $\mathbb{R}^n$, for any $1\le j\le n-1$,

$$\boxed{\mu_j} = \max_{\substack{U\subseteq\mathbb{R}^{n-1}\\ \dim U=j}}\left(\min_{\substack{y\in U\\ ||y||=1}}(By)\cdot y\right) = \max_{\substack{U\subseteq\mathbb{R}^{n-1}\times\{0\}\subseteq\mathbb{R}^n\\ \dim U=j}}\left(\min_{\substack{x\in U\\ ||x||=1}}(Ax)\cdot x\right)$$

(in other words, we thought of everything with an appended 0). But this is like the max-min term for λ_j of the matrix A, except with a smaller set of possibilities, so we can bound this as

$$\le \max_{\substack{U\subseteq\mathbb{R}^{n}\\ \dim U=j}}\left(\min_{\substack{x\in U\\ ||x||=1}}(Ax)\cdot x\right) = \boxed{\lambda_j}.$$

Finally, we also have (again by Theorem 64 on the matrix B) that

$$\boxed{\mu_j} = \min_{\substack{U\subseteq\mathbb{R}^{n-1}\\ \dim U=(n-1)-j+1}}\left(\max_{\substack{y\in U\\ ||y||=1}}(By)\cdot y\right) = \min_{\substack{U\subseteq\mathbb{R}^{n-1}\times\{0\}\subseteq\mathbb{R}^n\\ \dim U=n-(j+1)+1}}\left(\max_{\substack{x\in U\\ ||x||=1}}(Ax)\cdot x\right),$$

and again we're minimizing over only a subset of the subspaces in the min-max principle, so we can do better by minimizing over all subspaces:

$$\ge \min_{\substack{U\subseteq\mathbb{R}^{n}\\ \dim U=n-(j+1)+1}}\left(\max_{\substack{x\in U\\ ||x||=1}}(Ax)\cdot x\right) = \boxed{\lambda_{j+1}},$$

and chaining together all of these inequalities completing the proof. □

We'll state and prove the sensitivity conjecture next lecture, and the idea is that we'll want to repeatedly use the Cauchy interlacing theorem. So here's what we arrive at when we do that:

Corollary 66

Let A be a real symmetric $n\times n$ matrix, and let $I\subseteq\{1,\cdots,n\}$ be a subset of size $|I|=m$. Let B be the symmetric $m\times m$ submatrix obtained by only picking the rows and columns indexed within I. If $\lambda_1\ge\cdots\ge\lambda_n$ are the eigenvalues of A, and $\mu_1\ge\cdots\ge\mu_m$ are the eigenvalues of B, then for any $1\le j\le m$,

$$\lambda_j\ge\mu_j\ge\lambda_{j+n-m}.$$

In other words, we delete $(n-m)$ rows and columns from A, and each time the second inequality has index increased by one. (So we don't get the strictly interlacing pattern that we do in the Cauchy interlacing result, but we can still bound the eigenvalues of B in terms of the eigenvalues of A.) n n

13 March 17, 2022

We discussed Cauchy's interlace theorem last time, which explains that the eigenvalues of an $(n-1)\times(n-1)$ principal submatrix of a real symmetric $n\times n$ matrix interlace the eigenvalues of the original matrix. This result can then be extended to arbitrary-size $m\times m$ submatrices – we find that if the original eigenvalues are $\lambda_1\ge\cdots\ge\lambda_n$, and the submatrix eigenvalues are $\mu_1\ge\cdots\ge\mu_m$, then (we don't have strict interlacing in the same way anymore, but) $\lambda_j\ge\mu_j\ge\lambda_{j+n-m}$ for all $1\le j\le m$.

Today, we'll be using this to prove a more exciting result, the **sensitivity conjecture** (proved very recently, in 2019, by Hao Huang). This conjecture was posed in the 1990s, discussing sensitivity of Boolean functions, and it is important for applications in computer science, but we won't talk too much about the history of the problem here because of time constraints. (The introduction of Huang's paper does a good job giving background here if we're curious.) It turns out that the problem is equivalent to a problem about induced subgraphs of a hypercube:

Definition 67

Let Q^n be the n-dimensional hypercube graph, which has vertex set $\{0,1\}^n$ and has two vertices connected if they differ in exactly one coordinate (that is, their Hamming distance is 1).

For example, Q^2 looks like a square, and Q^3 looks like the wire outline of a cube.

Theorem 68 (Huang, 2019)

For any positive integer n, any $(2^{n-1}+1)$-vertex induced subgraph of the n-dimensional hypercube Q^n has maximum degree at least $\sqrt{n}$.

In other words, if we pick slightly more than half of the vertices of the hypercube, at least one of the vertices in that set would be connected to at least $\sqrt{n}$ of the others in that set. And we need to take more than half the vertices, because if we only take 2^{n-1} vertices, we can take all vertices with even sum of coordinates (in other words, "checkerboard 2-color" the hypercube graph), and then all vertices in the induced subgraph have degree 0. So somehow adding one more vertex makes a significant difference to this picture!

Fact 69

Many people had tried proving this result because of its importance, but the previously known best bound was $\left(\frac{1}{2} - o(1)\right) \log_2 n$ (proved in 1989). And in fact, there are constructions where the maximum degree is $\lceil \sqrt{n} \rceil$ (we'll see this on our homework), so this bound is exactly tight.

This proof uses spectral graph theory, but it uses a trick – we have to take a **signed adjacency matrix** in which some of the edges correspond to 1s and others correspond to -1s.

Lemma 70

Let G be an m-vertex graph, and let A be a symmetric $m \times m$ matrix with entries in $\{-1,0,1\}$ and rows and columns indexed by vertices $V(G)$. Suppose that $A_{u,v} \in \{-1,1\}$ if u and v are adjacent, and $A_{u,v} = 0$ otherwise (including the diagonals). If $\lambda_1(A) \geq \cdots \geq \lambda_m(A)$ are the eigenvalues of A, then $\lambda_1(A)$ is at most the maximum degree of G.

In other words, we take the adjacency matrix of G, and we place signs on the entries in any way as long as A stays symmetric. And the proof of this result is very similar to when we proved that a d-regular graph has a maximum eigenvalue of d:

Proof of lemma. Suppose we have an eigenvector $x = (x_v)_{v \in V(G)} \in \mathbb{R}^{V(G)}$ (coordinates also indexed by vertices) of our matrix A with eigenvalue $\lambda_1(A)$. Pick a vertex $u \in V(G)$ such that $|x_u|$ is maximal – we know that $|x_u| > 0$ (because otherwise x would be the zero vector). Then the uth coordinate of the eigenvalue equation $Ax = \lambda_1(A)x$ reads

$$\sum_{v \in V(G)} A_{u,v} x_v = \lambda_1(A) x_u,$$

and now the left-hand side is just a sum over the vertices v adjacent to u (because otherwise $A_{u,v} = 0$). Taking absolute values of both sides, we find

$$\left|\sum_{v\sim u} A_{u,v}x_v\right| = |\lambda_1(A)||x_u|,$$

and now by the triangle inequality and the fact that $A_{u,v} = \pm 1$ for all $v \sim u$, we find

$$|\lambda_1(A)||x_u| \le \sum_{v\sim u} |x_v| \le \deg(u)|x_u|.$$

Dividing through by $|x_u|$ gives us $|\lambda_1(A)| \le \deg(u)$, which is at most the maximum degree of G. Thus $\lambda_1(A)$ is indeed at most the max degree of G as well. □

If we look at this proof, we can see that the result would also hold if we allowed $A_{u,v} \in [-1, 1]$ for all $u \sim v$, but we only need the setting of the lemma for our signed adjacency matrices. And we'll now see how the insight for choosing the correct signed adjacency matrix in Huang's proof:

Proof of Theorem 68. Let A_n be the $2^n \times 2^n$ matrix defined recursively via the block recurrence

$$A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad A_n = \begin{bmatrix} A_{n-1} & I_{2^{n-1}} \\ I_{2^{n-1}} & -A_{n-1} \end{bmatrix},$$

where $I_{2^{n-1}}$ is the $2^{n-1} \times 2^{n-1}$ identity matrix (with 1s on the diagonal and 0s off the diagonal). We can check that A_n is a real symmetric matrix by induction, because the two I_n blocks flip to each other and $\pm A_{n-1}$ are both symmetric.

We claim that A_n is a signed adjacency matrix for Q^n – in other words, A_n has entries in $\{-1, 0, 1\}$, and if we replace all -1s with 1s, we get the adjacency matrix of Q^n (for some ordering of the vertices). The first part (entries all in $\{-1, 0, 1\}$ is clear by induction, and the second part follows by the following argument: order the vertices lexicographically (in other words, order them like the natural numbers when we encode them as binary strings), and let A_n^* be the adjacency matrix of Q^n with that ordering. Then $A_1^* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = A_1$ (because Q^1 is just a line segment connecting two vertices), and furthermore we have the same recurrence without the negative sign:

$$A_n^* = \begin{bmatrix} A_{n-1}^* & I_{2^{n-1}} \\ I_{2^{n-1}} & A_{n-1}^* \end{bmatrix}.$$

This is because the top left block comes from just looking at the "lower level" of the hypercube (all vertices with first coordinate 0), which has adjacency given by the $(n-1)$-hypercube graph Q^{n-1}. Similarly, the bottom right block comes from the "bottom level" of the hypercube (with first coordinate 1). And from there, the edges between the top and bottom level can only occur if we take the corresponding vertices on the top and bottom, since we can only have one coordinate of difference. So if we forget all of the signs in A_n, the signs in the recurrence relation don't matter anymore and we recover A_n^*.

So now if we want to apply Lemma 70, we need the eigenvalues of the matrix A_n. But that turns out to not be too bad:

Lemma 71

For all n, we have $A_n^2 = nI_{2^n}$.

Proof of lemma. We proceed again by induction: for $n = 1$, we have $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 = I_{2^1}$. Now for the inductive step, notice that by block multiplication,

$$A_n^2 = \begin{bmatrix} A_{n-1} & I_{2^{n-1}} \\ I_{2^{n-1}} & -A_{n-1} \end{bmatrix} \begin{bmatrix} A_{n-1} & I_{2^{n-1}} \\ I_{2^{n-1}} & -A_{n-1} \end{bmatrix} = \begin{bmatrix} A_{n-1}^2 + I_{2^{n-1}}^2 & A_{n-1} - A_{n-1} \\ A_{n-1} - A_{n-1} & I_{2^{n-1}}^2 + A_{n-1}^2 \end{bmatrix}$$

but now this simplifies nicely to

$$= \begin{bmatrix} (n-1)I_{2^{n-1}} + I_{2^{n-1}} & 0 \\ 0 & (n-1)I_{2^{n-1}} + I_{2^{n-1}} \end{bmatrix} = nI_{2^n},$$

as desired. □

This means that A_n^2 has all eigenvalues n, so A_n must have all eigenvalues $\pm\sqrt{n}$ – specifically, because the trace of A_n is zero (all diagonal entries are zero, or because the trace is $\text{tr}(A_{n-1}) - \text{tr}(A_{n-1}) = 0$ from the recursive formula), half of the eigenvalues of A_n are $\sqrt{n}$, and the other half are $-\sqrt{n}$. So if we tried applying Lemma 70 right now to A_n, we'd find that the maximum degree of Q^n is at least $\sqrt{n}$, which isn't helpful because the maximum degree is n.

But now if we take any subset $H \subseteq \{0,1\}^n$ of size $2^{n-1}+1$ of the vertices of our hypercube Q^n, we can let $G = Q^n[H]$ be the induced subgraph of Q^n induced by H. Remembering that A_n is a signed adjacency matrix for Q^n if we label the rows and columns in lexicographic order, we know that $(A_n)_{u,v}$ is in $\{1,-1\}$ if u and v are adjacent and 0 otherwise, so **the $(2^{n-1}+1) \times (2^{n-1}+1)$ submatrix B obtained by taking the rows and columns indexed by elements in H is a signed adjacency matrix for G**, and it's also symmetric. So by the Cauchy interlace theorem, the top eigenvalue μ_1 of B satisfies

$$\mu_1 \ge \lambda_{1+2^n-(2^{n-1}+1)} = \lambda_{2^{n-1}} = \sqrt{n},$$

so by Lemma 70 the maximum degree of B is at least $\sqrt{n}$, as desired. □

We can in fact see in the proof that if we only took a 2^{n-1}-vertex induced subgraph, the entire proof breaks down because the $(2^{n-1}+1)$th eigenvalue of A_n flips over to $-\sqrt{n}$ and we are no longer able to say anything useful! So it's pretty magical that everything works out so nicely (and in fact that the bound is sharp) – the key was finding a signed matrix A_n that manages to have a large 2^{n-1}th eigenvalue.

This is all we'll say about spectral graph theory for now – our next topic will be the **combinatorial nullstellensatz** (a return to polynomials), which is not a particularly difficult statement on its own but turns out to be very useful in applications. We won't state the result until next time, instead just stating a lemma that will be helpful (which is actually sort of morally equivalent to the combinatorial nullstellensatz):

Lemma 72

Let $\mathbb{F}$ be a field, and let $P \in \mathbb{F}[x_1, \cdots, x_n]$ be an n-variable polynomial such that the degree of P in x_i is at most t_i for each $1 \le i \le n$. Also let $S_i \subseteq \mathbb{F}$ be a subset of size $|S_i| = t_i + 1$ for each i. Then if $P(s_1, \cdots, s_n) = 0$ for all $(s_1, \cdots, s_n) \in S_1 \times \cdots \times S_n$, then P is the zero polynomial.

This is essentially a generalization of the result that a one-variable polynomial of degree t with at least $(t+1)$ roots is the zero polynomial, stated with more variables. In fact, we proved a special case of this for the finite field Kakeya problem, with the field being $\mathbb{F}_{q-1}$, all $t_i = q-1$, and all sets equal to $\mathbb{F}_q$. When we did the proof on our homework, we did a more indirect proof, constructing a map between polynomials and functions, showing that it's surjective and thus bijective by dimension counting, and deducing injectivity from that. We'll present a more direct method of proof this time:

Proof. As mentioned, the $n = 1$ case is showing that a nonzero one-variable polynomial of degree t has at most t zeros, which is true. For larger n, we use induction: consider $P(x_1, \cdots, x_n)$ as a polynomial in x_n, and write it as

$$P(x_1, \cdots, x_n) = Q_{t_n}(x_1, \cdots, x_{n-1})x_n^{t_n} + Q_{t_n-1}(x_1, \cdots, x_{n-1})x_n^{t_n-1} + \cdots + Q_0(x_1, \cdots, x_{n-1}).$$

Then for any fixed $(s_1, \cdots, s_{n-1})$, we have a degree-t_n polynomial $P(s_1, \cdots, s_{n-1}, x_n)$ in the variable t_n, which has $(t_n + 1)$ roots and thus must evaluate to zero for all x_n, meaning that all coefficients are zero:

$$Q_{t_n}(s_1, \cdots, s_{n-1}) = \cdots = Q_0(s_1, \cdots, s_{n-1}) = 0 \quad \forall (s_1, \cdots, s_{n-1}) \in S_1 \times \cdots \times S_{n-1}.$$

But each of these polynomials Q_i has degree at most $t_i + 1$ in each variable x_i, so by inductive hypothesis, each Q_i is actually the zero polynomial, and thus plugging this back in gives us $P = 0$, completing the proof. □

14 March 29, 2022

Last lecture (before break), we stated and proved a lemma which we'll now be able to use to (relatively easily) prove the combinatorial nullstellensatz. That lemma essentially stated that if a polynomial $P \in \mathbb{F}[x_1, \cdots, x_n]$ has degree at most t_i in each x_i, and we have sets S_i of size $t_i + 1$ such that $P(s_1, \cdots, s_n) = 0$ for all $s_i \in S_i$, then P must be the zero polynomial. The next result will look pretty similar but slightly strange:

Theorem 73 (Combinatorial Nullstellensatz (Alon))
Let $\mathbb{F}$ be a field, and let $P \in \mathbb{F}[x_1, \cdots, x_n]$ be an n-variable polynomial of degree d. Suppose that some monomial $x_1^{t_1} \cdots x_n^{t_n}$ is a monomial of maximum degree $t_1 + \cdots + t_n = d$ appearing in P with a nonzero coefficient. Let $S_1, \cdots, S_n \subseteq \mathbb{F}$ such that $|s_i| > t_i$ for all i. Then there is some $(s_1, \cdots, s_n) \in S_1 \times \cdots \times S_n$ such that $P(s_1, \cdots, s_n) \neq 0$.

(The word "nullstellensatz" means "theorem about zeros of a polynomial," though we should notice that this result is telling us that **not all** of the points in $S_1 \times \cdots \times S_n$ are zeros. And in algebraic geometry, there's a deeper theorem called Hilbert's Nullstellensatz, and after the proof we'll explain how we can rewrite this theorem in a similar way to that one.)

Proof. Without loss of generality we can assume that $|S_i| = t_i + 1$ for all i (because decreasing the size of the sets only makes the theorem harder to prove). Throughout this proof, we'll fix the field $\mathbb{F}$, as well as the exponents $t_1, \cdots, t_n$ and sets $S_1, \cdots, S_n$, but we'll be modifying the polynomial P until we can apply the lemma from last time. (Right now, there might be other monomials of degree d, so it is not necessarily true that the degree of P in the variable x_i is at most t_i.)

The idea is to subtract polynomials that vanish on all of $S_1 \times \cdots \times S_n$ from P, because making those modifications doesn't change the evaluation $P(s_1, \cdots, s_n)$. Some simple polynomial "building blocks" that do the job are

$$g_i(x_1, \cdots, x_n) = \prod_{s \in S_i} (x_i - s)$$

for each i. This indeed vanishes on all of $S_1 \times \cdots \times S_n$ because one of the terms in the product is zero, and g_i has degree $t_i + 1$. We'll now subtract multiples of these g_is from P, and the idea is that whenever we have a monomial with some x_i having degree at least $t_i + 1$, we can subtract off an appropriate multiple of g_i to get rid of it, and then repeat this process until all of the offenders are gone.

But it's generally a bit annoying to prove that some process terminates, and the easiest way is to consider some invariant property that decreases when we repeat our subtraction process. And whenever we have a proof of that sort, **the following strategy is usually best to make the proof short**. Suppose we have a counterexample $P \in \mathbb{F}[x_1, \cdots, x_n]$ which contradicts our theorem, meaning that (remembering $t_1, \cdots, t_n$ are all fixed) $x_1^{t_1} \cdots x_n^{t_n}$ appears with a nonzero coefficient but $P(s_1, \cdots, s_n) = 0$ for all $s_i \in S_i$. **Choose the counterexample to be minimal** using the following metric: suppose that any monomial of degree ℓ that appears in P contributes a weight 2^ℓ to P (so in particular, P always has a contribution of weight 2^d from $x_1^{t_1} \cdots x_n^{t_n}$ – we ignore the coefficient), and let the weight of P be the sum of all weights from its monomials. We'll then pick P to have the minimum weight. (This is possible because the weight is always a nonnegative integer and is always finite for a fixed d.)

Since P must contain the monomial $x_1^{t_1} \cdots x_n^{t_n}$, it cannot be the zero polynomial. So the conclusion of the lemma we proved last time is false, meaning that one of the assumptions does not hold. That's only possible if the degree assumption is not satisfied, meaning that there is some $1 \le i \le n$ such that the degree of P is at least $t_i + 1$ in x_i. Without loss of generality, take $i = n$ by reindexing for notational convenience. We then know that there is some nonzero monomial $cx_1^{a_1} \cdots x_n^{a_n}$ that appears in P, such that $c \neq 0$ (nonzero coefficient), $a_n \ge t_n + 1$, and $a_1 + \cdots + a_n \le d$. We'll now get rid of this polynomial by subtracting a multiple of g_n: define

$$P^*(x_1, \cdots, x_n) = P(x_1, \cdots, x_n) - cx_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n - t_n - 1} g_n(x_1, \cdots, x_n).$$

Because g_n and P both vanish on all of $S_1, \cdots, S_n$, so does P^*. Also, the second term has degree $a_1 + \cdots + a_{n-1} + a_n - t_n - 1 + (t_n + 1) = d$, so P^* is a difference of two terms of degree d and thus has degree at most d. (And notice that we have used the fact that $a_n \ge t_n + 1$ to have a well-defined polynomial here.) Furthermore, $x_1^{t_1} \cdots x_n^{t_n}$ still appears with a nonzero coefficient in P^* – it did so in P, and the only monomial of top degree d in the second term is $x_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n - t_n - 1} x_n^{t_n + 1}$ (grabbing the top monomial from g_n), which is $x_1^{a_1} \cdots x_n^{a_n} \neq x_1^{t_1} \cdots x_n^{t_n}$ because $a_n > t_n$. So P^* also satisfies the **same counterexample conditions** as P.

But now we claim that P^* has lower weight than P. Indeed, the only terms that differ between the two are

$$cx_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n - t_n - 1} g_n(x_1, \cdots, x_n) = cx_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n - t_n - 1} \left(x_n^{t_n + 1} + b_{t_n} x_n^{t_n} + \cdots + b_1 x_n + b_0 \right),$$

where the coefficients $b_0, \cdots, b_{t_n}$ are specified by the definition of g_n but aren't important, and this expands out to (coefficients in parentheses)

$$= (c)x_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n} + (b_{t_n} c)x_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n - 1} + \cdots + (b_0 c)x_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n - t_n - 1}.$$

But the first term here cancels out exactly with the $cx_1^{a_1} \cdots x_{n-1}^{a_{n-1}} x_n^{a_n}$ which we assumed was in P, so the 2^d weight contribution disappears in P^*. For all of the other terms, the weight might be increased or decreased, but the maximum possible weight contribution we can have is if we gain back a $2^{d-1} + 2^{d-2} + \cdots < 2^d$. So overall, P^*'s weight is smaller than P, contradicting the minimality assumption. Thus we could not have had our counterexample in the first place, and P must take on a nonzero value on some point of $S_1 \times \cdots \times S_n$. □

(This proof could have been rephrased without writing a counterexample by noting that the weight of P gets smaller by at least 1 each time, and this might be a more natural way to think about the proof. But it's easier to write things down in this "minimal counterexample" way.)

Fact 74

Another way to state the combinatorial nullstellensatz is that any polynomial $P \in \mathbb{F}[x_1, \cdots, x_n]$ which vanishes on all of $S_1 \times \cdots \times S_n$ must be of the form $P = h_1 g_1 + \cdots + h_n g_n$, where g_is are as in the proof and h_is are polynomials of degree at most $\deg P - \deg g_i$. (This follows by thinking about the reduction resulting in the zero polynomial.) And this is similar to Hilbert's Nullstellensatz, which states that over an algebraically closed field and with **any** polynomials of the form $g_1, \cdots, g_n$, if P vanishes on the common zeros of $g_1, \cdots, g_n$, then $P^k = h_1 g_1 + \cdots + h_n g_n$ for some $k, h_1, \cdots, h_n$.

We're now ready to see some combinatorial applications. The following result was posed by Komjath and solved by Alon and Füredi before the combinatorial nullstellensatz was stated, but the proof ended up being one of the motivating reasons for that formulation:

Theorem 75

Let $H_1, \cdots, H_m$ be (affine, $(n-1)$-dimensional) hyperplanes in $\mathbb{R}^n$, such that all but one vertex of the hypercube $\{0,1\}^n$ lies in $H_1 \cup \cdots \cup H_m$ (but the last vertex does not). Then $m \geq n$.

(Notice that we would be able to take $m = 2$ if we didn't have the constraint that one of the vertices does not lie in the union by using the hyperplanes $x_1 = 0$ and $x_1 = 1$, and that wouldn't be very interesting.) To get equality in the result, we can use the hyperplanes $x_1 = 1, x_2 = 1, \cdots, x_n = 1$, which covers all vertices except $(0, \cdots, 0)$, or similarly we can use the hyperplanes $x_1 + \cdots x_n = 1, 2, \cdots, n$.

Proof. Without loss of generality, assume that the origin $(0, \cdots, 0)$ is not covered, and suppose for contradiction that there is a collection of hyperplanes $m < n$. Each hyperplane H_i is described by a linear equation of the form

$$h_i(x_1, \cdots, x_n) = a_{i,1}x_1 + \cdots + a_{i,n}x_n - b_i = 0,$$

To get the union of the hyperplanes, we can multiply the h_is together: if h_i vanishes precisely on H_i for all i, then $h_1 h_2 \cdots h_m$ vanishes precisely on $H_1 \cup \cdots \cup H_m$, which contains all vertices of the hypercube except the origin. (In fact, $h_1 h_2 \cdots h_m(0, \cdots, 0) = (-b_1)(-b_2)\cdots(-b_m)$, and we'll call this nonzero value c.)

We're not ready to apply the combinatorial nullstellensatz yet, but we can consider the polynomial

$$P(x_1, \cdots, x_n) = h_1(x_1, \cdots, x_n) \cdots h_m(x_1, \cdots, x_n) - c(1 - x_1) \cdots (1 - x_n).$$

This second term vanishes at all other points of the hypercube except the origin, and because of our choice of c it also vanishes at $(0, \cdots, 0)$. So now we can check the conditions of the combinatorial nullstellensatz – the degree of P is n, because there is an $x_1 \cdots x_n$ monomial in the second term but the first term only has degree $m < n$. Furthermore, setting all $t_i = 1$ and all $S_i = \{0, 1\}$ (having size $2 > 1$), notice that $x_1^{t_1} \cdots x_n^{t_n}$ appears in P with a nonzero coefficient, in fact $(-1)^{n+1}c$, because again the first term $h_1 \cdots h_m$ cannot contribute to a degree-n monomial. So all conditions are satisfied, meaning that there should be some point in $S_1 \times \cdots \times S_n$ (the hypercube) where P is nonzero. This is a contradiction, so our initial assumption that $m < n$ must be incorrect (and $m \geq n$ is required). □

15 March 31, 2022

As a reminder to the graduate students in the class, there is a vote next Monday and Tuesday about whether a union will be formed. Professor Goemans has already sent the math department an email of links to resources in favor and opposed to the union, but what's important is for everyone who is eligible to vote. (Also, our third homework assignment is due today.)

We'll continue to talk about some applications of the combinatorial nullstellensatz today. This first result was initially proved in 1913, and it's about additive operations on sets:

Theorem 76 (Cauchy-Davenport)
Let p be a prime, and let A, B be nonempty subsets of $\mathbb{F}_p$. Let $A + B$ denote the set $\{a + b : a \in A, b \in B\}$. Then

$$|A + B| \geq \min(p, |A| + |B| - 1).$$

If we look at the set of integers instead of $\mathbb{F}_p$, it's relatively easy to show that $|A + B| \geq |A| + |B| - 1$ (by ordering the sets and constructing a set of increasing elements of $A + B$). But that argument doesn't work in $\mathbb{F}_p$ because of "wrap-around effects." It turns out this result is very tight – for example, we can achieve equality by having $A = \{1, \cdots, |A|\}$ and $B = \{1, \cdots, |B|\}$ as long as $|A| + |B| \leq p + 1$. And of course, we must always have $|A + B| \leq p$ (because $\mathbb{F}_p$ only has p elements).

Proof. Since the result involves a funny-looking minimum, we'll divide the situation into two cases. In **case 1**, suppose $|A| + |B| - 1 \geq p$, and we want to prove that $|A + B| \geq p$ – in other words, we must show that $A + B = \mathbb{F}_p$, or equivalently that any $x \in \mathbb{F}_p$ is the sum of an element in A and an element in B. Indeed, the sets A and $x - B = \{x - b : b \in B\}$ have a total of $|A| + |B| > p$ elements, so they must intersect and we have $a = x - b$ for some $a \in A, b \in B$, as desired.

Now for **case 2**, suppose $|A| + |B| - 1 \leq p$ (this overlaps with case 1 slightly, but it's good enough for our proof). Suppose for the sake of contradiction that $|A + B| \leq |A| + |B| - 2 \leq p - 1$. We wish to apply the combinatorial nullstellensatz, and we'll do this by finding a polynomial and a grid on which it vanishes – it makes sense to have that grid be $A \times B$, and the polynomial $P \in \mathbb{F}_p[x, y]$ we'll use is

$$P(x, y) = \prod_{c \in A+B} (x + y - c).$$

The idea is that for any $x \in A$ and $y \in B$, $x + y$ will be some element $c \in A + B$, so $P(a, b) = 0$ for all $a \in A$ and $b \in B$. We'll now check the conditions of the combinatorial nullstellensatz: the degree of P is $|A + B|$, and we should look at the monomial $x^{t_1} y^{t_2}$ such that $t_1 + t_2 = |A + B|$ (maximal degree) and $|A| > t_1, |B| > t_2$. Since we know that $|A + B| \leq |A| + |B| - 2$, we are indeed always able to find t_1 and t_2 that satisfy those conditions (though note that depending on the actual degree of P, we might need to choose different values of t_1, t_2 – we can't always pick $t_1 = |A| - 1, t_2 = |B| - 1$, for example, because $|A + B|$ might be strictly less than $|A| + |B| - 2$).

Now the coefficient of $x^{t_1} y^{t_2}$ in P is just $\binom{t_1+t_2}{t_1}$, because we can't take any c's for a maximal-degree monomial and thus the coefficient is the same as for $(x + y)^{|A+B|} = (x + y)^{t_1+t_2}$. And that coefficient **is** nonzero, because we're working in $\mathbb{F}_p$ but $t_1 + t_2 \leq p - 1$ so there's no terms divisible by p here. So all conditions of the combinatorial nullstellensatz hold with $S_1 = A, S_2 = B$, and there should be a point $(a, b) \in A \times B$ with $P(a, b) \neq 0$, a contradiction. This finishes the proof. □

Theorem 77 (Chevalley-Warning)
Let p be a prime, and let $Q_1, \cdots, Q_m \in \mathbb{F}_p[x_1, \cdots, x_n]$ be m polynomials in n variables such that $\deg(Q_1) + \cdots + \deg(Q_m) < n$. Then the number of common zeros of the polynomials $Q_1, \cdots, Q_m$ in $\mathbb{F}_p^n$ is divisible by p.

(Chevalley originally proved a weaker result, which is that whenever we have a common zero, there must be another one as well.)

Proof. Let A be the set of common zeros:

$$A = \{(a_1, \cdots, a_n) \in \mathbb{F}_p^n : Q_i(a_1, \cdots, a_n) = 0 \quad \forall 1 \le i \le m\}.$$

We must show that p divides $|A|$. Suppose for the sake of contradiction that this is not the case – it makes sense to construct another polynomial which encodes the property of being a common zero. Taking the product of the Q_is would give us the union of the zeros, but if we want the intersection of the zeros we want to do something like

$$P^*(x_1, \cdots, x_n) = \prod_{i=1}^{m} \left(1 - Q_i(x_1, \cdots, x_n)^{p-1}\right).$$

This polynomial ends up being 1 if $(a_1, \cdots, a_n)$ is a common zero of all polynomials (because it's a product of $(1-0)$s), and otherwise it ends up being 0 because $a^{p-1} = 1$ for any nonzero a. Furthermore, $\deg P^* \le (p-1)(\deg(Q_1) + \deg(Q_2) + \cdots + \deg(Q_m)) < (p-1)n$ by assumption. Applying the combinatorial nullstellensatz to this would not be very useful, because we already know when P^* is nonzero and in fact already know there is a common zero (because we assumed $|A|$ is not divisible by p). So instead, we'll consider the polynomial which actually subtracts off the 1s at each common zero $(a_1, \cdots, a_n)$:

$$P(x_1, \cdots, x_n) = P^*(x_1, \cdots, x_n) - \sum_{(a_1, \cdots, a_n) \in A} \left(1 - (x_1 - a_1)^{p-1}\right) \cdots \left(1 - (x_n - a_n)^{p-1}\right).$$

Indeed, this second term is what we constructed on our homework: this polynomial vanishes whenever $x_i \ne a_i$ for any i because our product gets a $(1-1)$ factor, and otherwise we get $(1-0)\cdots(1-0) = 1$. So we've now constructed a polynomial which vanishes on all of $\mathbb{F}_p^n$.

To apply the combinatorial nullstellensatz, we now need to extract more properties of P. We know that $\deg P \le (p-1)n$, because P^* has degree less than $(p-1)n$ and each term in the sum has degree $(p-1)n$ as well. We'll now take $t_1 = \cdots = t_n = p-1$ and let each S_i be all of $\mathbb{F}_p$. The monomial $x_1^{p-1} \cdots x_n^{p-1}$ now has coefficient $|A|(-1)^{n+1}$, because there are no contributions from P^* and each term in the sum gives us a $(-1)^{n+1}$ coefficient. This is nonzero by assumption because $|A| \ne 0$ in $\mathbb{F}_p$ (in fact, this tells us that the degree of P is actually $(p-1)n$, **so it's valid to take** $t_1 = \cdots = t_n = p-1$), so the conditions of the combinatorial nullstellensatz hold and we have that P is nonzero somewhere. This is a contradiction, and thus $|A|$ must be divisible by p. □

Remark 78. *An alternative proof that doesn't use the combinatorial nullstellensatz is to notice (after constructing P^* above) that*

$$|A| = \sum_{(a_1, \cdots, a_n) \in \mathbb{F}_p^n} P^*(a_1, \cdots, a_n).$$

But if we look at any individual monomial that might show up in P^*, adding it over all of $\mathbb{F}_p^n$ will always give us 0 because of the degree condition (details left to us). Thus $|A| = 0$ in $\mathbb{F}_p$.

We'll next discuss the "Erdös-Ginzburg-Ziv constant of $\mathbb{F}_p$" – this name will make sense later:

Theorem 79 (Erdös-Ginzburg-Ziv (1961))

Let p be a prime. Then any sequence of elements of $\mathbb{F}_p$ of length $(2p-1)$ contains a subsequence of length p of sum zero.

(This result might look strange, but zero-sum-subsequence problems are a whole field in combinatorics.) The value $(2p-1)$ above is tight – indeed, consider $(p-1)$ 1s followed by $(p-1)$ 0s. Then any length p subsequence must have anywhere between 1 and $(p-1)$ 1s and thus does not sum to 0 in $\mathbb{F}_p$.

Proof. Write the sequence as $a_1, \cdots, a_{2p-1}$. We wish to pick "indicator" variables $x_1, \cdots, x_{2p-1} \in \{0,1\}$ such that $x_1 + \cdots + x_{2p-1} = p$ **in** $\mathbb{Z}$, **not in** $\mathbb{F}_p$, and $x_1a_1 + \cdots + x_{2p-1}a_{2p-1} = 0$ in $\mathbb{F}_p$. We'll basically use these x_is as our variables and $\{0,1\}$ as our sets S_i, and towards applying the combinatorial nullstellensatz we'll consider the polynomial $(x_1a_1 + \cdots + x_{2p-1}a_{2p-1})^{p-1} - 1$. We know that this polynomial is zero whenever our condition $x_1a_1 + \cdots + x_{2p-1}a_{2p-1} = 0$ is not satisfied and -1 whenever it is. We also need to require that $x_1 + \cdots + x_{2p-1} = p \in \mathbb{Z}$, which is a harder condition to encode, but we can still say that $(x_1 + \cdots + x_{2p-1})^{p-1} - 1$ is nonzero whenever $x_1 + \cdots + x_{2p-1} \neq 0$ in $\mathbb{F}_p$. So putting this together, we can consider

$$P^*(x_1, \cdots, x_{2p-1}) = \left[(x_1a_1 + \cdots + x_{2p-1}a_{2p-1})^{p-1} - 1\right]\left[(x_1 + \cdots + x_{2p-1})^{p-1} - 1\right].$$

This polynomial indeed vanishes unless $x_1a_1 + \cdots + x_{2p-1}a_{2p-1} = 0$ and $x_1 + \cdots + x_{2p-1} = 0$ in $\mathbb{F}_p$, and it has degree at most $2(p-1)$. But since $0 \le x_1 + \cdots + x_{2p-1} \le 2p-1$, those conditions hold either if we have a valid subsequence of length p **or** if all x_is are zero. And indeed, $P^*(0, \cdots, 0) = (-1)(-1) = 1$, so we just need to do something extra to P^* to allow us to apply the combinatorial nullstellensatz. Specifically, define

$$P(x_1, \cdots, x_{2p-1}) = P^*(x_1, \cdots, x_{2p-1}) - \prod_{i=1}^{2p-1}(1 - x_i).$$

This is an important detail: normally, we'd use $\prod_{i=1}^{2p-1}(1 - x_i^{p-1})$, but that makes the degree of P gigantic. But because we're only interested in the polynomial vanishing on $\{0,1\}^{2p-1}$ eventually, we can reduce the degree dramatically. So now $P(0, \cdots, 0) = 0$, and on the rest of $\{0,1\}^{2p-1}$ P only vanishes when we have a valid subsequence corresponding to $x_1, \cdots, x_{2p-1}$.

So applying the combinatorial nullstellensatz to the polynomial P, setting all $t_i = 1$, and using the sets $S_i = \{0,1\}$, there must be a point $(x_1, \cdots, x_{2p-1}) \in \{0,1\}^{2p-1}$ with $P(x_1, \cdots, x_{2p-1}) \neq 0$. We've calculated that this point is not the origin, so because $\prod_{i=1}^{2p-1}(1 - x_i)$ vanishes on the rest of $\{0,1\}^{2p-1}$, this means $P^*(x_1, \cdots, x_{2p-1}) \neq 0$. That happens only if $x_1a_1 + \cdots + x_{2p-1}a_{2p-1} = 0$ and $x_1 + \cdots + x_{2p-1} = 0$ in $\mathbb{F}_p$, which implies that $x_1 + \cdots + x_{2p-1} = p$ in $\mathbb{Z}$ because it's not zero and is at most $(2p-1)$. Thus we have a characteristic vector that encodes our subsequence, and we're done. □

16 April 5, 2022

Last week, we proved a statement about subsequences of $(2p-1)$-element sequences in $\mathbb{F}_p$ (really, $\mathbb{Z}_p$). We can restate this to a more general problem:

Problem 80

Given positive integers m and n, what is the minimum s such that among any s points in $\mathbb{Z}^n$, one can find m points whose centroid is also a lattice point in $\mathbb{Z}^n$?

We can notice that an s always exists – indeed, asking for m points with a lattice point centroid is the same as asking for the sum of coordinates (in each direction) to be a multiple of m. So we can "project down to $\mathbb{Z}_m^n$" (just consider the residue classes of coordinates mod m), and now if we have m copies of any of those m^n residue classes of coordinates, we definitely will have enough points. So by the pigeonhole principle, $s = (m-1)m^n + 1$ is definitely enough, and we can restate the problem above equivalently with this language:

Problem 81

Given positive integers m and n, what is the minimum s such that for any sequence of s (not necessarily distinct) elements in $\mathbb{Z}_m^n$, we have a subsequence of length m whose elements sum to 0 in $\mathbb{Z}_m^n$? (This s is known as the **Erdös-Ginzburg-Ziv constant** and is denoted $s(\mathbb{Z}_m^n)$.)

This is the version of the problem we solved last time with $m = p$, $n = 1$, and we showed that $s(\mathbb{Z}_p) = 2p-1$ in that case using the combinatorial nullstellensatz. And the pigeonhole argument above proves that $s(\mathbb{Z}_m^n) \le (m-1)m^n + 1$ for any m, n, and in fact for $m = 2$ this bound is tight (that is, $s(\mathbb{Z}_2^n) = 2^n + 1$). This is because in characteristic 2, the only way to get a subsequence $\{a_1, a_2\}$ of length 2 with $a_1 + a_2 = 0$ is if $a_1 = a_2$, so the pigeonhole principle is the only "limiting factor."

Fact 82

Determining these Erdös-Ginzburg-Ziv constants is difficult in general – Erdös-Ginzburg-Ziv proved in 1961 that $s(\mathbb{Z}_m) = 2m-1$ for general m (and $n = 1$), and Reiher proved in 2003 that $s(\mathbb{Z}_m^2) = 4m-3$. Also, for $m = 2^k$ (and all n), we know $s(\mathbb{Z}_{2^k}^n) = (2^k - 1)2^n + 1$. But that's all of the "infinite families" for which we know the answer.

On the other hand, we have the general lower bound $s(\mathbb{Z}_m^n) \ge (m-1)2^n + 1$, because we can always take $(m-1)$ copies of the set of all $\{0,1\}^n$ vectors and that will not be enough to get a subsequence summing to zero.

To shed light on how these results relate to each other, we have the following:

Lemma 83

For all positive integers m, n, k, we have $s(\mathbb{Z}_{mk}^n) \le k(s(\mathbb{Z}_m^n) - 1) + s(\mathbb{Z}_k^n)$.

In other words, it suffices to really consider the case where m is prime, because we can then repeatedly apply the lemma. For example, plugging in $s(\mathbb{Z}_p) = 2p - 1$ gives us the general $s(\mathbb{Z}_m) = 2m - 1$ for all m. (We'll see the proof of this on our last homework assignment for the class.)

Following these arguments, we've thus basically proved all of the statements in Fact 82 except that $s(\mathbb{Z}_m^2) = 4m-3$. To show the lower bound for that, we're basically using the $s(\mathbb{Z}_m^n)$ lower bound above: consider the sequence of length $(4m-4)$ containing $(m-1)$ copies of $(0,0)$, $(m-1)$ copies of $(0,1)$, $(m-1)$ copies of $(1,0)$, and $(m-1)$ copies of $(1,1)$. Then the only way to get a sum of m things to be zero is if we have a sum (in $\mathbb{Z}$) of either 0 or m in each coordinate, but that's not possible because we only have $(m-1)$ copies of each of $(0,0)$ and $(1,1)$.

For the upper bound, it turns out that Lemma 83 allows us to just prove the statement for m prime. Indeed, induct on the number of prime factors of m, and for the inductive step if $s(\mathbb{Z}_m^2) = 4m-3$ and $s(\mathbb{Z}_k^2) = 4k-3$, then

$$s(\mathbb{Z}_{mk}^2) \le k((4m-3)-1) + (4k-3) = 4mk - 3.$$

So now let's state the upper bound in a way that doesn't require the definition of $s(\mathbb{Z}_m^n)$. In particular, to avoid multisets, we'll switch to $\mathbb{Z}^2$ so that we can just use different points to represent the same class in $\mathbb{Z}_p^2$:

Theorem 84 (Reiher, 2003)

We have $s(\mathbb{Z}_p^2) \le 4p-3$ for all primes p. In other words, for any prime $p \ge 3$ and any subset $X \subseteq \mathbb{Z}^2$ of size $|X| = 4p-3$, there is some subset $\Gamma \subseteq X$ of size $|\Gamma| = p$ such that the sum of both coordinates of the points in Γ is divisible by p.

(The $p = 2$ case has already been discussed by Pigeonhole, so we don't need to worry about it.) We'll first do an easier case where we're given an extra point to work with:

Start of proof for $|X| = 4p-2$. We'll use the following notation: if we have a subset $Y \subseteq \mathbb{Z}^2$ and an integer $k \ge 0$, we'll let $(k|Y)$ denote the **number of** subsets I of Y of size k, such that the sum of the points in I has both coordinates divisible by p. For example, $(0|Y) = 1$ because the empty subset has sum of points 0 in each coordinate, and $(k|Y) = 0$ if $k > |Y|$. Our goal is then to prove that $(p|X) > 0$ if $|X| = 4p-3$ (but for this proof just $|X| = 4p-2$).

Lemma 85

For any subset $Y \subseteq \mathbb{Z}^2$ of size at least $3p-2$, we have

$$1 - (p|Y) + (2p|Y) - (3p|Y) + \cdots \equiv 0 \bmod p.$$

Note that the first term in this sum is just $(0|Y)$, and the sum will terminate because $(k|Y) = 0$ for large enough k. And the reason we're proving such a weird statement is that it's hard to say anything about a particular $(k|Y)$, because the combinatorial nullstellensatz can't distinguish between sizes of sets mod p. (We got around this with $s(\mathbb{Z}_p) = 2p-1$ because luckily $(2p-1)$ is smaller than $2p$, but that was basically a lucky coincidence.)

Proof of lemma. Write the points in Y as $\{(a_1, b_1), \cdots, (a_n, b_n)\} \subseteq \mathbb{Z}^2$ (where $|Y| = n \ge 3p-2$ by assumption). We're interested in subsets I of size divisible by p with coordinate sums of elements in I both zero. Because p is an odd prime, notice that the sum $\sum_I (-1)^{|I|}$ over all such valid subsets I is exactly the left-hand side of the lemma, because all size $p, 3p, 5p, \cdots$ subsets contribute a -1 to the sum, and all size $0, 2p, 4p, \cdots$ subsets contribute a $+1$ to the sum, just like in the original alternating sum. Thus we must show that $\sum_I (-1)^{|I|}$ is a multiple of p.

Indeed, any subset I of Y corresponds to an indicator vector $(x_1, \cdots, x_n) \in \{0,1\}^n$, where x_i is 1 if and only if I contains the point (a_i, b_i). With that notation, we're asking for our subset to satisfy

$$x_1a_1 + \cdots + x_na_n = 0, \quad x_1b_1 + \cdots + x_nb_n = 0, \quad x_1 + \cdots + x_n = 0$$

all in $\mathbb{F}_p$, so in other words the thing we are trying to compute on the left-hand side (and show is 0 in $\mathbb{F}_p$) is

$$\sum_{\substack{x=(x_1,\cdots,x_n)\in\{0,1\}^n \\ \text{satisfying above constraints}}} (-1)^{|x|},$$

where $|x| = x_1 + \cdots + x_n$ denotes the number of ones in the indicator vector $(x_1, \cdots, x_n)$. To encode this sum more explicitly, we want a polynomial $Q(x_1, \cdots, x_n)$ which encodes whether those constraints are satisfied, and like last lecture, we'll use

$$Q(x_1, \cdots, x_n) = (1 - (x_1a_1 + \cdots + x_na_n)^{p-1})(1 - (x_1b_1 + \cdots + x_nb_n)^{p-1})(1 - (x_1 + \cdots + x_n)^{p-1}).$$

This polynomial evaluates to 1 in $\mathbb{F}_p$ if all conditions are satisfied, and otherwise it evaluates to 0 by Fermat's little theorem (because one of the x_i sums will be nonzero, so its $(p-1)$th power will be 1). But we don't want to use the combinatorial nullstellensatz on Q just yet, because that's not going to give us useful statements of the sort we want. Instead, we should use a polynomial Q where we're "surprised" if it doesn't vanish, and much like last time we subtract the indicator function of each of the points $(x_1, \cdots, x_n)$ corresponding to a subset I satisfying the conditions. For any $y = (y_1, \cdots, y_n)$ satisfying our conditions, we can define

$$\delta_y(x_1, \cdots, x_n) = \text{the unique multilinear polynomial in the variables } x_1, \cdots, x_n$$
$$\text{such that for all } (x_1, \cdots, x_n) \in \{0,1\}^n, \delta_y(x_1, \cdots, x_n) = \begin{cases} 1 & x = y, \\ 0 & \text{otherwise.} \end{cases}$$

By "multilinear" here, we mean that the polynomial δ_y only contains monomials where each x_i only has an exponent at most 1, and this is fine because we're only going to need our modified Q polynomial to vanish on $\{0,1\}^n$. For example, if $y = (1, \cdots, 1)$, then $\delta_y(1, \cdots, 1) = x_1 \cdots x_n$, and if $y = (0, \cdots, 0)$, then $\delta_y(0, \cdots, 0) = (1 - x_1) \cdots (1 - x_n)$. More generally, δ_y is a product with factor x_i if $y_i = 1$ and $(1 - x_i)$ otherwise. Notice that $x_1 \cdots x_n$ has coefficient $(-1)^{n-|y|} = (-1)^n(-1)^{|y|}$ in δ_y (because we get a negative sign from each zero in y). With this, we can finally construct the polynomial

$$P(x_1, \cdots, x_n) = Q(x_1, \cdots, x_n) - \sum_{\substack{y=(y_1,\cdots,y_n)\in\{0,1\}^n \\ \text{satisfying above constraints}}} \delta_y(x_1, \cdots, x_n).$$

This polynomial vanishes on all of $\{0,1\}^n$, and we want to look at the coefficient of $x_1 \cdots x_n$ in P. There's no contribution to that coefficient from Q, because Q has degree $(3p - 3) < n$ by assumption of the lemma. Since we get a $(-1)^n(-1)^{|y|}$ from each y satisfying the condition, the total coefficient $x_1 \cdots x_n$ is

$$-\sum_{\substack{y=(y_1,\cdots,y_n)\in\{0,1\}^n \\ \text{satisfying above constraints}}} (-1)^n(-1)^{|y|}.$$

Now **assume for the sake of contradiction** that this coefficient is nonzero. Then P is a polynomial of degree n (because each term in the definition of P has degree at most n and the $x_1 \cdots x_n$ coefficient is nonzero), and applying the combinatorial nullstellensatz with all $t_i = 1$ and all sets $S_i = \{0,1\}$, we find that P must be nonzero somewhere on $\{0,1\}^n$, which is a contradiction. Thus we must indeed have this sum be zero, meaning that

$$-\sum_{\substack{y=(y_1,\cdots,y_n)\in\{0,1\}^n \\ \text{satisfying above constraints}}} (-1)^n(-1)^{|y|} = 0 \text{ in } \mathbb{F}_p.$$

Dividing by $(-1)^{n+1}$ gives us the desired result. □

In other words, this lemma shows us that the terms $1-(p|Y)+(2p|Y)-\cdots$ together form the vanishing coefficient $x_1 \cdots x_n$ in the polynomial P.

Lemma 86

Suppose $Y \subseteq \mathbb{Z}^2$ has size $|Y| = 3p$, and the sum of **all** points in Y has both coordinates divisible by p, then $(p|Y) > 0$.

In other words, if the sum of the coordinates is zero for all of Y, then that property is also true for some p-element subset of Y.

Proof of lemma. Suppose for the sake of contradiction that $(p|Y) = 0$. If Y' is any subset of size $3p - 1$, then $(p|Y') = 0$ as well (because if we can't find any subsets of Y that satisfy the coordinate sum condition, we definitely won't find any in Y'). Then by Lemma 85, we have $1-(p|Y')+(2p|Y') \equiv 0 \bmod p$ because Y' only has size $(3p-1)$, and this means that $(2p|Y') \equiv -1 \bmod p$. In particular, there is some subset S of size $2p$ with coordinate sum zero, but then this is a contradiction because we can take the complement S^c which will be a set of size p with coordinate sum zero as well (since the total coordinate sum is zero). □

Next lecture, we'll actually prove the $(4p-2)$ bound using these lemmas, but the point is to next go further by looking at $(3p-2)$-element subsets of our set X of size $(4p-2)$.

□

17 April 7, 2022

We'll continue the proof from last time, working towards determining $s(\mathbb{Z}_m^n)$ for $n = 2$. In particular, last time, we reduced the problem to the case where m is an odd prime, and our goal became showing that for a subset $X \subseteq \mathbb{Z}^2$ of size $(4p-3)$, there is a subset $I \subseteq X$ of size p with sum of both coordinates divisible by p.

Continuation of proof for $|X| = 4p-2$. Recall our notation from last time: for a subset $Y \subseteq \mathbb{Z}^2$, we let $(k|Y)$ be the number of size-k subsets with sum of both coordinates divisible by p. (So our eventual goal is to show that whenever X has size at least $(4p-3)$ (but $(4p-2)$ in this proof), we have $(p|X) > 0$.) We proved last time that whenever Y is a subset of size at least $(3p-2)$, we have $1-(p|Y)+(2p|Y)-(3p|Y)+\cdots \equiv 0 \bmod p$ (using the combinatorial nullstellensatz), and from that we proved that for any subset Y of size $3p$ with sum of coordinates of all points equal to zero, we must have $(p|Y) > 0$.

We can now do the main proof: for the sake of contradiction, suppose $(p|X) = 0$. Then for any subset Y of X, we also have $(p|Y) = 0$ (because if Y had a subset with coordinate sum zero in each direction, then X would also have that subset). We now claim that $(3p|X) = 0$ as well – indeed, if there were a subset Y of X of size $3p$ with sum of the coordinates in Y divisible by p, then we must have $(p|Y) > 0$ by our above lemmas, which is a contradiction. So in the equation

$$1-(p|X)+(2p|X)-(3p|X) \equiv 0 \bmod p$$

that we derived in general last time (there are no more terms because $4p$ is already larger than $|X|$), we can plug in $(p|X) = 0$ and $(3p|X) = 0$ to get

$$(2p|X) \equiv -1 \bmod p.$$

But we can do even more from here: if we consider any subset $Y \subseteq X$ of size $(3p-2)$, we have $(p|Y) = 0$, so the analogous equation also tells us that

$$1-(p|Y)+(2p|Y) \equiv 0 \bmod p \implies (2p|Y) \equiv -1 \pmod{p}.$$

We'll now do a double-counting argument – **we count the number of triples** (I, Y, X) modulo p with $I \subseteq Y \subseteq X$ and I, Y, X of size $(2p), (3p-2)$, and $(4p-2)$, respectively, such that I has a divisibility condition (but Y doesn't have any such requirement). On the one hand, this count is (**modulo** p)

$$(2p|X)\binom{2p-2}{p-2} \equiv -\binom{2p-2}{p-2},$$

because we pick one of the subsets of size $(2p)$ with the divisibility condition, and then to get $Y \supseteq I$ we pick $(p-2)$ of the remaining $(2p-2)$ elements. (We then use the fact that $(2p|X) \equiv -1 \bmod p$ from above.) On the other hand, this count is also (still modulo p)

$$\binom{4p-2}{3p-2}(2p|Y) \equiv -\binom{4p-2}{3p-2},$$

because we pick any subset Y of size $(3p-2)$ and then count how many $(2p)$-element subsets will work for that Y (and it's always -1 modulo p for any Y). So these two quantities should be equal modulo p, meaning that $\binom{2p-2}{p} \equiv \binom{4p-2}{p}$. Expanding out the binomial coefficients, this means

$$\frac{(2p-2)(2p-3)\cdots(p)(p-1)}{p!} \equiv \frac{(4p-2)\cdots(3p)(3p-1)}{p!} \bmod p.$$

It's dangerous to divide by p here when we have a congruence statement mod p, so let's cancel those out:

$$\frac{(2p-2)(2p-3)\dots(p+1)(p-1)}{(p-1)!} \equiv 3\frac{(4p-2)(4p-3)\cdots(3p+1)(3p-1)}{(p-1)!} \bmod p.$$

But now both fractions go through all nonzero residue classes on the numerator and denominator, so they completely cancel out, and we're left with

$$1 \equiv 3 \bmod p,$$

which is a contradiction because p is an odd prime. Thus we must have $(p|X) \neq 0$ at the start, as desired. □

Remark 87. *In this proof, even though the congruence statements only required that $(p|X) \equiv 0 \bmod p$, we* **do** *actually use the fact that we're assuming (for contradiction) that $(p|X)$ is exactly zero, because we need $(p|Y) = 0$ for subsets Y of X as well.*

We can notice that this proof does not work for $|X| = 4p-3$, because if we try the same size $|Y| = 3p-2$, the binomial coefficients become $\binom{2p-3}{p-2} = \binom{2p-3}{p-1}$ and $\binom{4p-3}{3p-2} = \binom{4p-3}{p-1}$, and those both evaluate to 0 mod p and we don't get a contradiction. On the other hand, if we try to use $|Y| = 3p-3$ instead, our lemma requiring $|Y| \geq 3p-2$ from the combinatorial nullstellensatz no longer holds. But we're now ready to consider the more difficult case of $|X| = 4p-3$, and we'll establish a stronger lemma:

Lemma 88

For any subset $Y \subseteq \mathbb{Z}^2$ of size $|Y| \geq 3p-3$, we have

$$1-(p-1|Y)-(p|Y)+(2p-1|Y)+(2p|Y)-(3p-1|Y)-(3p|Y)+\cdots \equiv 0 \bmod p.$$

In other words, we take integers that are either 0 or -1 mod p in a grouped alternating sum.

Proof idea of lemma. With the original lemma, we encoded the divisibility condition with a degree $(p-1)$ polynomial. But now that we are allowing for our subsets of Y to have two different size residue classes, we only need a degree $(p-2)$ polynomial to encode that: specifically, change the polynomial Q to

$$Q(x_1,\cdots,x_n) = (1-(x_1a_1+\cdots+x_na_n)^{p-1})(1-(x_1b_1+\cdots+x_nb_n)^{p-1})\prod_{s=1}^{p-2}((x_1+x_2+\cdots+x_n-s)$$

(where we should remember that $Y = \{(a_1,b_1),\cdots,(a_n,b_n)\}$). So this third factor vanishes as long as the size of our subset, encoded by $|x|$, is not 0 or -1 mod p, which is what the lemma statement suggests. The total degree of Q is then $(3p-4)$, and then the rest of the proof (subtracting the δ_ys and using the combinatorial nullstellensatz) then follows identically to the simpler version with $|Y| = 3p-2$. □

Proof of Reiher's result for $|X| = 4p-3$. Again, suppose for the sake of contradiction that $(p|X) = 0$, so that we still have $(p|Y) = 0$ for all subsets Y of X, and it's still true that $(3p|X) = 0$ from the other lemma. Now by Lemma 88, for any subset $Y \subseteq X$ of size $|Y| = 3p-3$, we have

$$1 - (p-1)|Y) - (p|Y) + (2p-1)|Y) + (2p|Y) \equiv 1 - (p-1|Y) + (2p-1|Y) + (2p|Y) \equiv 0 \bmod p.$$

Last time, we just had $1 - (2p|Y) \equiv 0 \bmod p$, and we added up $(2p|Y)$ across all possible subsets of Y. We'll use that same idea here: summing this congruence relation over all subsets Y of size $3p-3$ of X gives us

$$\binom{4p-3}{3p-3} - \sum_Y (p-1|Y) + \sum_Y (2p-1|Y) + \sum_Y (2p|Y) \equiv 0 \bmod p.$$

But now each of these terms can be written in terms of X: any subset of size $(p-1)$ of sum 0 in each coordinate shows up in $\binom{|X|-(p-1)}{p} = \binom{3p-2}{p}$ different terms of Y (we choose which p elements not in our $(p-1)$-element subset make up the complement $X \setminus Y$), and similar arguments work for the other terms:

$$\binom{4p-3}{3p-3} - (p-1|X)\binom{3p-2}{p} + (2p-1|X)\binom{2p-2}{p} + (2p|X)\binom{2p-3}{p} \equiv 0 \bmod p.$$

Rewriting the first term as $\binom{4p-3}{p}$, we can basically use the same trick as we did in the last proof: for example, $\binom{4p-3}{p} \equiv 3 \bmod p$ because the $3p$ factor cancels with the p when we expand out the factorials, and then the numerator and denominator contain all nonzero residue classes. This argument eventually gives us

$$\boxed{3 - 2(p-1|X) + (2p-1|X) + (2p|X) \equiv 0 \bmod p}.$$

But we also know that for **every** subset Y of size $3p-2$ or $3p-1$, our original lemma gives us

$$1 - (p|Y) + (2p|Y) \equiv 0 \bmod p \implies (2p|Y) \equiv -1 \bmod p.$$

So if we now do **another** double-counting argument (which might seem unmotivated), where we count the number of partitions $X = A \cup B \cup C$ such that $|A| = p-1, |B| = p-2, |C| = 2p$, and A and C have the coordinate sum divisibility constraint (corresponding to the terms $(p-1|X)$ and $(2p|X)$). To count this, we can first count by picking A or by picking B. If we start by choosing A, we have (**modulo** p) $-(p-1|X)$ ways to form a partition, because the complement of A is always of size $(3p-2)$, and then within that set we always have $(2p|A^c) \equiv -1 \bmod p$. On the other hand, if we start by choosing B, we have $-(3p-1|X)$ ways modulo p, because we need B's complement to have the divisibility constraint, and then after that there are always $(2p|B^c) \equiv -1 \bmod p$ ways to choose C. Thus we find that

$$-(p-1|X) \equiv -(3p-1|X) \bmod p \implies (p-1|X) \equiv (3p-1|X) \bmod p.$$

So now if we directly apply Lemma 88 to X, we find (using $(p|X) = (3p|X) = 0$, and also using the relations above)

$$\begin{aligned} \boxed{0} &\equiv 1 - (p-1|X) - (p|X) + (2p-1|X) + (2p|X) - (3p-1|X) - (3p|X) \bmod p \\ &\equiv 1 - (p-1|X) - 0 + (2p-1|X) + (2p|X) - (p-1|X) - 0 \bmod p \\ &\boxed{\equiv 1 - 2(p-1|X) + (2p-1|X) + (2p|X) \bmod p}. \end{aligned}$$

But again the boxed conditions tell us that $1 \equiv 3 \bmod p$, so we again get our contradiction. □

18 April 12, 2022

We've been discussing Erdös-Ginzburg-Ziv constants for the last class or two – recall that determining $s(\mathbb{Z}_m^n)$, the minimum number of points in $\mathbb{Z}^n$ required to find m points whose average is a lattice point, is an open question in general but known for some special cases (like $n = 1, 2$ or $m = 2^k$). We've mentioned previously that we have some general bounds

$$(m-1)2^n + 1 \le s(\mathbb{Z}_m^n) \le (m-1)m^n + 1, \quad s(\mathbb{Z}_{mk}^n) \le k(s(\mathbb{Z}_m^n) - 1) + s(\mathbb{Z}_k^n),$$

which in particular tell us that it's mostly interesting to consider the cases where m is prime.

Fact 89

It turns out that for any fixed dimension n, $s(\mathbb{Z}_m^n)$ grows linearly with m. A linear lower bound is easy from our bound $(m-1)2^n+1 \le s(\mathbb{Z}_m^n)$ above, but the first upper bound shown was that $s(\mathbb{Z}_m^n) \le (cn\log_2 n)^n \cdot m$ for some absolute constant c (interesting in the range where m is very large and n is held fixed), shown by Alon and Dubiner in 1995. The result has recently been improved – it's been shown by Zakharov in 2020 that $s(\mathbb{Z}_p^n) \le 4^n \cdot p$ if p is a prime that is sufficiently large with respect to n. (But because those primes can be arbitrarily large, we can't use the recursive bound $s(\mathbb{Z}_{mk}^n) \le k(s(\mathbb{Z}_m^n) - 1) + s(\mathbb{Z}_k^n)$ to get a clean bound for all m.) We were originally going to discuss the proof of $s(\mathbb{Z}_m^n) \le (cn\log_2 n)^n \cdot m$, but we'll move along to stay back on track with the syllabus.

We can also consider the opposite situation, where we fix m and let the dimension n get large (in which the problem is less well understood). Assuming that m is prime (again, this is the most interesting case because of the recursive bound), we can think about small values of m. We already know that $s(\mathbb{F}_2^n) = 2^n + 1$, and it turns out that computing $s(\mathbb{Z}_3^n)$ is equivalent to the next topic of our class, the famous **cap-set problem** and the slice rank polynomial method.

Problem 90 (Cap-set problem)

What is the largest size of a subset of $\mathbb{F}_3^n$ without three points on an affine line?

In particular, notice that every affine line in $\mathbb{F}_3^n$ contains exactly three points, so our question is equivalently "how large can we make a subset $A \subseteq \mathbb{F}_3^n$ where A does not contain an entire line," and in affine geometry "cap-set" refers to exactly a set A of this form. An easy lower bound to establish is 2^n by using $A = \{0, 1\}^n$ – to see why this set does not contain three points on a line, we can verify that x, y, z are on a line if and only if $(x - y) - (y - z) = x + y + z = 0$ (specifically only in characteristic 3), and if we have three distinct points in $\{0, 1\}^n$ their sum will not be zero. And we also have the upper bound $\frac{3^n+1}{2}$, because if we look at all of the lines through a given point in the set A, those lines partition the remaining points in $\mathbb{F}_3^n$ into pairs, and we can only have one of each pair. But even though we have exponential lower and upper bounds for this problem, we still don't actually have a satisfying upper bound because the total set of the size is 3^n:

Fact 91

The best known lower bound is approximately $(2.2174\cdots)^n$ (due to Edel in 2004), and the first upper bound that tends to a negligible fraction of the whole set was $O(3^n/n)$ (due to Meshulam in 1995). This bound was later improved to $3^n/(n^{1+c})$ for some small constant $c > 0$ (due to Bateman and Katz in 2012 – even this tiny improvement was considered a big breakthrough). And most recently, Ellenberg and Gijswijt showed in 2017 that we can actually get a bound of 2.756^n, the first exponential base upper bound smaller than 3.

Ellenberg and Gijswijt's result can be stated more generally as follows:

Theorem 92 (Ellenberg–Gijswijt (2017))

Let $p \geq 3$, and let $A \subseteq \mathbb{F}_p^n$ be a set not containing any nontrivial 3-term arithmetic progressions (meaning that there are no distinct $x, y, z \in \mathbb{F}_p^n$ such that $x - 2y + z = 0$). Then $|A| \leq (\Gamma_p)^n$, where

$$\Gamma_p = \min_{0<t<1} \frac{1 + t + \cdots + t^{p-1}}{t^{(p-1)/3}}$$

is a constant depending only on p and strictly smaller than p.

(Notice that this fraction $\frac{1+t+\cdots+t^{p-1}}{t^{(p-1)/3}}$ approaches infinity at $t \to 0$, approaches p as $t \to 1$, and is continuous on $(0, 1]$. But because the derivative is positive at $t = 1$, we must have a minimum achieved strictly between 0 and 1 – the calculus might be a little annoying to check but is not too important.) The proof of this result very importantly relies on another result by Croot, Lev, and Pach for $\mathbb{Z}_4^n$, and Ellenberg and Gijswijt independently figured out how to modify that result for $\mathbb{F}_p^n$ (it was not immediately clear to them). And Tao later reformulated the problem to a more general version (this version is now called the **slice rank polynomial method**); that's the proof we'll be following in this class. All of these developments occurred within a month or so of each other!

We've seen various "polynomial methods" in this class (such as dimension counting, counting zeros with the joints and Kakeya problems, and the combinatorial nullstellensatz). This one is a new one and was novel enough that there were cool applications applying it in the months following those developments. To use it, we'll need to first define what "slice rank" means:

Definition 93

Let $\mathbb{F}$ be a field, A be a finite set (indexed in any way), and $k \geq 2$ be an integer. The **slice rank** of a function $f : A^k \to \mathbb{F}$ is defined in the following way: f has slice rank 1 if it can be written as $f(x_1, \cdots, x_k) = g(x_j)h(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k)$ for nonzero functions $g : A \to \mathbb{F}$ and $h : A^{k-1} \to \mathbb{F}$ and some $1 \leq j \leq k$. Then the slice rank of f is the minimum r such that f can be written as the sum of r slice rank 1 functions (with potentially different js).

Remark 94. *A function $f : A^k \to \mathbb{F}$ is also called a* ***k*-tensor** *– that word sometimes has scary connotations because of commutative algebra, but we can also think of it as a "hypermatrix" because, for example, a function $f : A^2 \to \mathbb{F}$ can be represented as an $|A|$ by $|A|$ matrix which just encodes the numbers $f(i, j)$.*

Example 95

The slice rank of a function $f : A^k \to \mathbb{F}$ is always at most $|A|$ (this is kind of like how an $n \times n$ matrix can only be of rank at most n). To see this, look at the "horizontal slices" of our hypercube A^k – letting δ_a be the indicator function for $x_1 = a$ (for any $a \in A$), we have

$$f = \sum_{a \in A} \delta_a(x_1) \cdot f(a, x_2, \cdots, x_k),$$

where each term in the product is of slice rank 1.

Note that the **slice rank is not the same as the ordinary tensor rank** (in which a rank 1 tensor looks like $g_1(x_1)g_2(x_2)\cdots g_k(x_k)$). This means that the slice rank is always smaller than the tensor rank, but the most general upper bound for the rank of a k-tensor is $|A|^k$. On the other hand, we can check that for $k = 2$, **the slice rank is the same as the ordinary rank of a matrix**. Remember that diagonal matrices with nonzero entries have rank n, we have an analogous result for our functions:

Lemma 96 (Tao)

Suppose $f : A^k \to \mathbb{F}$ is such that $f(x_1, \cdots, x_k) \neq 0$ if and only if $x_1 = \cdots = x_k$. Then the slice rank of f is $|A|$.

Proof. We induct on k. For $k = 2$, this result is equivalent to the fact that diagonal matrices with nonzero entries have full rank. For the inductive step, because the slice rank of any function is at most $|A|$, we can assume for the sake of contradiction that the slice rank is less than $|A|$. Then we can write f as a sum of fewer than $|A|$ slice rank 1 functions

$$\boxed{f(x_1, \cdots, x_k) = \sum_{\alpha \in M_1} g_\alpha(x_1) h_{\alpha(x_2, \cdots, x_k)} + \sum_{\alpha \in M_2} g_\alpha(x_2) h_\alpha(x_1, x_3, \cdots, x_k) + \cdots + \sum_{\alpha \in M_k} g_\alpha(x_k) h_\alpha(x_1, \cdots, x_{k-1})},$$

where M_i are disjoint index sets with $|M_1| + \cdots + |M_k| < |A|$ (in other words, each slice rank 1 function in the sum is of the form $g_\alpha(x_j) h_\alpha(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k)$ for some j, and so we index the functions of this form with this particular j by M_j). So now consider the space of all functions $\phi : A \to \mathbb{F}$ such that

$$\sum_{x \in A} \phi(x) g_\alpha(x) = 0 \quad \text{for all } \alpha \in M_k;$$

in other words, we want our functions ϕ to be orthogonal to all of the functions of the form $g_\alpha(x_k) h_\alpha(x_1, \cdots, x_{k-1})$ that show up in our sum for f. This space has dimension at least $|A| - |M_k|$ (we have at most $|M_k|$ linearly independent constraints on ϕ), so there is some subset A' of size at least $|A| - |M_k|$ and some ϕ such that $\phi(x)$ is nonzero for all $x \in A'$. (This is because $\sum_{x \in A} \phi(x) g_\alpha(x) = 0$ can be thought of as linear equations in the variables $\phi(x)$ – row reducing and taking A' to be the set of free variables, we can let ϕ be the function which takes value 1 on all free variables.) Now defining the function $f' : (A')^{k-1} \to \mathbb{F}$ via

$$f'(x_1, \cdots, x_{k-1}) = \sum_{x_k \in A} f(x_1, \cdots, x_k) \phi(x_k),$$

this function is indeed diagonal because f is diagonal, and it has nonzero entries on the diagonal because f is nonzero on the diagonal and $\phi(x_k) \neq 0$ on A'. Thus by the inductive hypothesis, the slice rank of A' is at least $|A| - |M_k|$. But on the other hand, multiplying the boxed equation by $\phi(x_k)$ and then summing over all $x_k \in A$ gives us

$$f'(x_1, \cdots, x_{k-1}) = \sum_{\alpha \in M_1} g_\alpha(x_1) \left(\sum_{x_k \in A} h_\alpha(x_2, \cdots, x_k) \phi(x_k) \right) + \cdots + \sum_{\alpha \in M_{k-1}} g_\alpha(x_{k-1}) \left(\sum_{x_k \in A} h_\alpha(x_1, \cdots, x_{k-2}, x_k) \phi(x_k) \right)$$

(the last term vanishes by the definition of ϕ). So the slice rank of f' is $|A'| \geq |A| - |M_k|$ but is also at most $|M_1| + \cdots + |M_{k-1}|$, giving us a contradiction because we assumed that $|M_1| + \cdots + |M_k| < |A|$. □

Next lecture, we'll see how we can use this lemma to prove Ellenberg and Gijswijt's result!

19 April 14, 2022

We'll be proving Ellenberg and Gijswijt's result today – recall that it states that a subset A of $\mathbb{F}_p^n$ that does not contain a nontrivial 3-term progression satisfies $|A| \leq (\Gamma_p)^n$, where $\Gamma_p = \min_{0<t<1} \frac{1+t+\cdots+t^{p-1}}{t^{(p-1)/3}}$. As we discussed last time, the special case $p = 3$ is the cap-set problem, and this result gives us an exponentially better bound than what was previously possible for the problem. And for some numerical reference, it turns out $0.8414p \le \Gamma_p \le 0.9184p$.

Fact 97

Recall that a "nontrivial three-term progression" means that we have three distinct elements x, y, z with $x-2y+z = 0$. But the "wrap-around" nature of $\mathbb{F}_p^n$ for restricting three-term progressions is very important here – there are subsets of $\{1, \cdots, N\}$ in $\mathbb{Z}$ without a three-term arithmetic progression of size at least $e^{-c\sqrt{\log N}}N$. So looking at these sets in comparison to the ground set, in the $\mathbb{F}_p^n$ case we know that $|A|$ must be at most size $(p^n)^{1-c_p}$ for some constant c_p, but in the $\mathbb{Z}$ case we can have $|A|$ of size $(N)^{1-o(1)}$.

We'll be using the slice-rank polynomial method to prove this result – recall that a function $f : A^k \to \mathbb{F}$ has slice rank 1 if we can write it as a product $f(x_1, \cdots, x_k) = g(x_j)h(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k)$ for some j, and generally a function's slice rank is the minimum number of slice rank 1 functions that must be added together to get f. We proved last time (by "horizontal slicing") that the slice rank is always at most $|A|$ and that it is exactly $|A|$ for a "diagonal" function f (in which the values $f(x_1, \cdots, x_k)$ are nonzero only when $x_1 = \cdots = x_k$).

Proof of Ellenberg–Gijswijt's result. Suppose we have such a set A. For any point $x \in A \subseteq \mathbb{F}_p^n$, we'll write $x = (x^{(1)}, \cdots, x^{(n)})$ with upper indices to avoid conflicting notation. We need to construct a "tensor" (function), and specifically we'll do so with our set A which contains no nontrivial three-term arithmetic progressions: define the function $f : A \times A \times A \to \mathbb{F}_p$ given by

$$f(x, y, z) = \prod_{i=1}^{n} \left((x^{(i)} - 2y^{(i)} + z^{(i)})^{p-1} - 1 \right).$$

Because x, y, z form an arithmetic progression (**in that order**) if and only if $x - 2y + z = 0$, this is the same as saying that $x^{(i)} - 2y^{(i)} + z^{(i)} = 0$ for all i, and then we're playing the usual game with the combinatorial nullstellensatz – if any of these coordinate calculations are nonzero, the expression will be 0, but if we do have an arithmetic progression it will be $(-1)^n$.

But by assumption A has no nontrivial 3-term arithmetic progressions, the only way for f to be nonzero is to have a 3-term arithmetic progression with a repeated term, and because $p \ge 3$ this only occurs if $x = y = z$ (we have to be careful here – something like $1, 0, 1$ would work in $p = 2$!). In fact, $f(x, y, z)$ is indeed a "diagonal" function of the sort we talked about, because $f(x, x, x) = (-1)^n$ for all $x \in A$ and otherwise $f(x, y, z) = 0$. Thus f has slice rank $|A|$ by our lemma.

It may not seem like we've made very much progress, but now the polynomial nature of f will give us an upper bound for the slice rank. Specifically, because f does not have a very high degree (at most $(p-1)n$), we will be able to write f explicitly as a sum of slice rank 1 functions, which will be how we get the bound on $|A|$. In particular, a degree $(p-1)n$ polynomial is a sum of monomials of degree at most $(p-1)n$ in $x^{(1)}, \cdots, x^{(n)}, y^{(1)}, \cdots, y^{(n)}, z^{(1)}, \cdots, z^{(n)}$, such that any individual variable has degree at most $(p-1)$. Each monomial has degree distributed between the x's, y's, and z's, so it must have degree at most $\frac{(p-1)n}{3}$ in either the x's, y's, or z's. Explicitly, the ones where the degree is at most $\frac{(p-1)n}{3}$ in the x's look like $(x^{(1)})^{d_1} \cdots (x^{(n)})^{d_n}$ times a monomial in y and z (where $d_1 + \cdots + d_n \le \frac{(p-1)n}{3}$), and then there are similar expressions for the groups for y's and z's. And the idea is that **we'll group (by the distributive law) all of the terms with the same d_is in** x, and also do the same grouping among the same d_is in y's and the same d_is in z's. This means that we can write f as a sum of functions of the form

$$(x^{(1)})^{d_1} \cdots (x^{(n)})^{d_n} \cdot (\text{polynomial}(y, z)),$$

plus also similar expressions $(y^{(1)})^{d_1} \cdots (y^{(n)})^{d_n} \cdot (\text{polynomial}(x,z))$ and $(z^{(1)})^{d_1} \cdots (z^{(n)})^{d_n} \cdot (\text{polynomial}(x,y))$. And the point is that these are **valid slice rank (at most, in case the polynomial is zero) 1 functions** – these terms are each a function of x times a polynomial in y and z, or a function of y times a polynomial in x and z, or a function of z times a polynomial in x and y. (Importantly, we couldn't have broken this up into $x^{(1)}$ times a polynomial in the other coordinates, because that wouldn't be separating out x from the other variables.) So overall, we find that the slice rank of f can be bounded as

$$|A| = \text{slice rank } f \le 3\left|\{(d_1, \cdots, d_n) \in \{0, \cdots, p-1\}^n : d_1 + \cdots + d_n \le \frac{(p-1)n}{3}\right|,$$

since we get a slice rank 1 polynomial for every valid set of exponents d_i. And now the rest is computation (we're going to lose a factor of about $\sqrt{n}$ to get to the nicer-looking $(\Gamma_p)^n$, but it generally isn't seen as very important) – we'll show that $|A| \le 3\Gamma_p^n$ with the following lemma:

Lemma 98

Pick a uniformly random n-tuple $(d_1, \cdots, d_n) \in \{0, \cdots, p-1\}^n$. Then the probability that $d_1 + \cdots + d_n \le \frac{(p-1)n}{3}$ is at most $(\Gamma_p/p)^n$.

Proof of lemma. Since Γ_p is the minimum of $\frac{1+t\cdots+t^{p-1}}{t^{(p-1)/3}}$ for $t \in (0,1)$, it suffices to prove that the probability is at most that fraction for any t. Indeed, fixing some t,

$$\mathbb{P}\left(d_1 + \cdots + d_n \le \frac{(p-1)n}{3}\right) = \mathbb{P}\left(t^{d_1+\cdots+d_n} \ge t^{(p-1)n/3}\right)$$

because $t < 1$, and then by Markov's inequality we can bound this by

$$\le \frac{\mathbb{E}[t^{d_1+\cdots+d_n}]}{t^{(p-1)n/3}} = \frac{\mathbb{E}[t^{d_1}]^n}{(t^{(p-1)/3})^n} = \left(\frac{(1+t+\cdots+t^{p-1})/p}{t^{(p-1)/3}}\right)^n = \left(\frac{1}{p}\frac{1+t\cdots+t^{p-1}}{t^{(p-1)/3}}\right)^n.$$

Since this holds for any $t \in (0,1)$, it holds for the minimum of all such t, which gives us

$$\mathbb{P}\left(d_1 + \cdots + d_n \le \frac{(p-1)n}{3}\right) \le \frac{1}{p^n}\Gamma_p^n,$$

as desired. □

In particular, this means that the fraction of all p^n possible n-tuples $(d_1, \cdots, d_n)$ that are valid is at most $\frac{(\Gamma_p)^n}{p^n}$, so there are at most $(\Gamma_p)^n$ terms for each of the x, y, and z slice rank 1 functions, giving us $|A| \le 3(\Gamma_p)^n$ overall. To remove the factor of 3, the idea is to now use the **power trick**: notice that if A is a 3-term progression-free subset, then A^m is also 3-term progression-free (where we're thinking of A^m as isomorphic to $\mathbb{F}_p^{nm}$, so we're avoiding 3-term arithmetic progressions for (nm)-dimensional vectors). Thus applying the bound we have already proved, we find that

$$|A|^m \le 3(\Gamma_p)^{nm} \implies |A| \le 3^{1/m}(\Gamma_p)^n,$$

and taking m arbitrarily large (taking the infimum) implies that $|A| \le (\Gamma_p)^n$, completing the proof. □

Fact 99

There is still a gap between the currently known lower and upper bound, and in particular it's not actually known whether this upper bound is tight. And even though the Markov bound in our lemma might look weak, it turns out it's not necessarily a bad bound. In particular, a generalization of this problem (tri-color sum-free theorem) is actually tight with its corresponding Γ_p^* – the argument there involves coupling the probability distribution with itself. (And there's a k-colored version of all of this too, in which the bound is also tight – this was shown in a paper coauthored by Professor Sauermann.)

Notice that Lemma 98 is essentially a Chernoff bound, because we expect $d_1 + \cdots + d_n$ to be $\frac{(p-1)n}{2}$ on average. So that is the point at which we saw why **we needed the polynomial f to be of low degree** – we needed it to have degree smaller than $\frac{3}{2}(p-1)n$. But that also explains why this proof method **does not** work for looking at 4-term progressions, which essentially require two equations in four variables (such as $x - 2y + z = 0, y - 2z + w = 0$). Then we'd have a polynomial $f(w, x, y, z)$ of degree $2(p-1)n$, but then the slice-rank splitting would have four different types and each having $d_1 + \cdots + d_n \le \frac{(p-1)n}{2}$, which is not good enough. More generally, m equations in k variables will only give a meaningful bound if $k \ge 2m+1$, so trying to do k-term arithmetic progressions will never work this way for $k \ge 4$.

To be more precise with all of this, suppose we have a single equation in $k \ge 3$ variables of the form $a_1x_1 + \cdots + a_kx_k = 0$ (which is the equation we're trying to avoid among points $x_i \in A$). In order to use this proof method to bound $|A|$, we must have $\boxed{a_1 + \cdots + a_k = 0}$ (otherwise the set $A \subseteq \mathbb{F}_p^n$ of all points with first coordinate 1 will not have any solutions to $a_1x_1 + \cdots + a_kx_k$, and it has size $\frac{1}{p}p^n$) – this comes up in the fact that our tensor $f : A^k \to \mathbb{F}$ must be diagonal and have **nonzero** entries on the diagonal. And there's another caveat as well – instead of requiring x, y, z to be all distinct, we must require them to not be all equal (so we have a stronger set of conditions):

Theorem 100 (Generalization of Ellenberg–Gijswijt)

Let p be a fixed prime, let $k \ge 3$, and let $a_1, \cdots, a_k \in \mathbb{F}_p \setminus \{0\}$ such that $a_1 + \cdots + a_k = 0$. Suppose $A \subseteq \mathbb{F}_p^n$ is a set such that there are no points $x_1, \cdots, x_k \in A$ with $a_1x_1 + \cdots + a_kx_k = 0$ with the x_is not all equal. Then $|A| \le (\Gamma_{p,k})^n$, where $\Gamma_{p,k} = \min_{0<t<1} \frac{1+t+\cdots+t^{p-1}}{t^{(p-1)/k}} < p$.

20 April 19, 2022

Last lecture, we mentioned that Ellenberg and Gijswijt's result about progression-free subsets of $\mathbb{F}_p^n$ (which we proved last time) generalizes a more general problem. Specifically, if $A \subseteq \mathbb{F}_p^n$ is some set where there is no set of not-all-equal elements $x_1, \cdots, x_k \in A$ satisfying $a_1x_1 + \cdots + a_kx_k = 0$ (where $k \ge 3$, a_i are all nonzero, and $a_1 + \cdots + a_k = 0$), then we can bound the size of A as $|A| \le (\Gamma_{p,k})^n$, where $\Gamma_{p,k} = \min_{0<t<1} \frac{1+t+\cdots+t^{p-1}}{t^{(p-1)/k}}$ is some constant independent of n and (importantly) strictly less than p. (So we have an exponentially strong bound on $|A|$ which in fact gets better for larger k.) Ellenberg and Gijswijt's theorem is this result with $p \ge 3$, $a_1 = 1$, $a_2 = -2$, $a_3 = 1$, and the way to prove the more general problem is with essentially the same proof as last time (using the slice rank polynomial method). And we can in fact further generalize this result to systems of equations by taking multiple sets of equations of the form $a_1x_1 + \cdots + a_kx_k = 0$, as long as the number of variables k is strictly more than twice the number of equations m; then, instead of $t^{(p-1)/k}$, we have $t^{m(p-1)/k}$ in the denominator for $\Gamma_{p,k}$.

Remark 101. *We mentioned that our proofs didn't tell us whether the exponent bases Γ_p and $\Gamma_{p,k}$ are tight but that* *there was a similar situation in which the exponent base is known to be tight. The actual setup for that is listed below, but we won't focus on this too much.*

Theorem 102 (Multi-colored sum-free theorem)

Fix a prime p and let $k \geq 3$. Consider a list of L k-tuples, indexed as $(y_{1,\ell}, \cdots, y_{k,\ell}) \in \mathbb{F}_p^n \times \cdots \times \mathbb{F}_p^n$ for $\ell \in \{1, \cdots, L\}$. Suppose that for any $\ell_1, \cdots, \ell_k \in \{1, \cdots, L\}$, we have $y_{1,\ell_1} + \cdots + y_{k,\ell_k} = 0$ if and only if $\ell_1 = \cdots = \ell_k$. Then $L \leq (\Gamma_{p,k})^n$ for the same $\Gamma_{p,k}$ as above, and in fact we have a tight bound (there is no better constant than $\Gamma_{p,k}$; a lower bound of $(\Gamma_{p,k})^{n-O(\sqrt{n})}$ is known).

In other words, we can write out our k-tuples, placing $(y_{1,1}, \cdots, y_{2,1}, \cdots, y_{k,1})$ in the first row, $(y_{1,2}, \cdots, y_{2,2}, \cdots, y_{k,2})$ in the second row, and so on. The result then says that the sum of any row is zero, but there's no other way to take one element of the first component, one element of the second component, and so on, and have them add to 0. (Think of each component of its own color.) And it turns out that to recover our original generalization, we correspond every $x \in A$ with the k-tuple $(a_1 x, \cdots, a_k x)$.

With that, we can return to the topic of Erdös-Ginzburg-Ziv constants from earlier. If we take $k = p$ and all $a_i = 1$ in our generalization of Ellenberg–Gjiswijt, we get the following result:

Theorem 103

Let $p \geq 5$ be a prime (the bound is otherwise not interesting), and let $A \subseteq \mathbb{F}_p^n$ be a subset that does not contain $x_1, \cdots, x_p \in A$, not all equal, with $x_1 + \cdots + x_p = 0$. Then $|A| \leq \Gamma_{p,p}^n \leq 4^n$.

(We also get a corresponding lower bound of 2^n from $\{0,1\}^n$ as usual.)

Proof. Given the work we've done already, we just need to verify that $\Gamma_{p,p} \leq 4$. Indeed,

$$\Gamma_{p,p} = \min_{0<t<1} \frac{1 + \cdots + t^{p-1}}{t^{(p-1)/p}} \leq \min_{0<t<1} \frac{1/(1-t)}{t} = 4$$

by replacing the numerator with the infinite series and noting that $t^{(p-1)/p} \geq t$ for $0 < t < 1$, and then finally noting that $t(1-t)$ has a maximum at $t = 1/2$. □

Recall that we defined $s(\mathbb{Z}_m^n)$ to be the smallest s such that any sequence of s elements in $\mathbb{Z}_m^n$ has a subsequence of length m whose elements sum to 0 in $\mathbb{Z}_m^n$. We mentioned that a motivating question was to understand the behavior of this constant when we fix m and take n to infinity, and in particular it suffices to understand the behavior when m is prime. We've now managed to almost convert our results back to that language. However, the issue with immediately generalizing is that in Theorem 103, we only assume that $x_1, \cdots, x_p$ are **all not equal**, while in Erdös-Ginzburg-Ziv we assume that they must **all be distinct**. (And the slice rank proof required our "diagonal" function to have extremely few nonzero entries, so that proof will not work directly here.)

Furthermore, in Erdös-Ginzburg-Ziv, it's okay to "repeat" elements of $\mathbb{F}_p^n$, while we haven't been allowing that in our generalization of Ellenberg–Gijswijt. But that's an easier complication to deal with. For that, let's define $s^*(\mathbb{Z}_p^n)$ to be the maximum size of a subset $A \subseteq \mathbb{F}_p^n$ in which we forbid solutions $x_1 + \cdots + x_p = 0$ with x_i all distinct (so it's like $s(\mathbb{Z}_p^n)$, but this time we allow repetition of x_is). Then we have

$$s^*(\mathbb{Z}_p^n) + 1 \leq s(\mathbb{Z}_p^n),$$

because the set A in the Erdös-Ginzburg-Ziv setting shows that we can't forbid a sum of p things adding to zero for $s(\mathbb{Z}_p^n)$ so the best value for $s^*(\mathbb{Z}_p^n)$ is at most $s(\mathbb{Z}_p^n)+1$, and we also have

$$s(\mathbb{Z}_p^n) \le (p-1)s^*(\mathbb{Z}_p^n)+1$$

because if we have a vector with p repetitions that gives us a subsequence and we're happy in the Erdös-Ginzburg-Ziv setting, and otherwise we have at least $s^*(\mathbb{Z}_p^n)+1$ different vectors and thus there are p of them that are distinct and sum to 0. Thus $s^*(\mathbb{Z}_p^n)$ and $s(\mathbb{Z}_p^n)$ differ by a factor of at most p, so up to a constant factor in the fixed p regime these are basically the same.

So putting this together, we've already proved that $s(\mathbb{Z}_2^n) = 2^n+1$, and from Ellenberg–Gjiswijt and the argument we just made, we have $s(\mathbb{Z}_3^n) \le 2(\Gamma_3)^n + 1 \le 2 \cdot 2.756^n + 1$. (We're lucky here because "x_1, x_2, x_3 sum to zero and are distinct" is the same as "x_1, x_2, x_3 sum to zero and are not all equal.") Now for a prime $p \ge 5$, we want the largest possible size of a subset $A \subseteq \mathbb{F}_p^n$ without distinct $x_1, \cdots, x_p \in A$ summing to zero. Here's the historical rundown: in 2017 (but published in 2020), Naslund proved that $|A| \le p \cdot 2^p \cdot (\Gamma_p)^n$ (remember we can think of Γ_p as roughly $0.9p$), by introducing the concept of **partition rank**. But then a better bound was proved, for which we can go over the proof now:

Theorem 104 (Fox–S. (2017))
If $A \subseteq \mathbb{F}_p^n$ does not contain p distinct vectors with $x_1, \cdots, x_p \in \mathbb{F}_p^n$ with $x_1 + \cdots + x_p = 0$, then $|A| \le 3(\Gamma_p)^n$. In other words, using the argument above, we have $s(\mathbb{Z}_p^n) \le 3p(\Gamma_p)^n$.

In fact, Professor Sauermann has more recently improved this bound to $|A| \le c_p \sqrt{p\Gamma_{p,p}}^n \le c_p(2\sqrt{p})^n$, which has base now much better than linear in p but still far away from the 4^n behavior in Theorem 103. And that bound generalizes to the multi-colored version too and is tight there.

Proof. The idea is to first ask "how often a given element of A can be the middle term of a 3-term arithmetic progression" and reduce to the generalization of Ellenberg-Gijswijt. In particular, if there is some $x \in A$ which appears as the middle term of $\frac{p-1}{2}$ 3-term arithmetic progressions in A, then the p vectors appearing in these progressions are p distinct vectors that add to $p \cdot x = 0$ (in particular the arithmetic progressions only overlap at x). So every $x \in A$ appears as the middle term of at most $\frac{p-3}{2}$ 3-term arithmetic progressions, so there are at most $|A| \cdot \frac{p-3}{2}$ nontrivial 3-APs in A. (As a note, for $p = 3$ any element can be the middle element of the progression, so for that case we can just fix a choice of middle element at the start.)

Now let H be a uniformly random affine hyperplane in $\mathbb{F}_p^n$, and let $X_1 = |A \cap H|$ and X_2 be the number of nontrivial 3-term arithmetic progressions in $A \cap H$. We have

$$\mathbb{E}[X_1] = \frac{1}{p}|A|$$

(because every point in $\mathbb{F}_p^n$ shows up in H with probability $\frac{1}{p}$), and we have

$$\mathbb{E}[X_2] = \frac{1}{p}\frac{p^{n-1}-1}{p^n-1} \cdot (\text{number of nontrivial 3-term arithmetic progressions in } A) \le \frac{1}{p^2} \cdot \left(\frac{p-3}{2}|A|\right) < \frac{1}{2p}|A|$$

(where we're finding the probability that two of the points in any given arithmetic progression are in A, at which point the third will also be). Thus by the triangle inequality, we have

$$\mathbb{E}[X_1 - X_2] > \frac{1}{p}|A| - \frac{1}{2p}|A|,$$

so there is some hyperplane H in which $X_1 - X_2 > \frac{1}{2p}|A|$. Now construct a subset B of $A \cap H$ in which we delete one point from each nontrivial 3-term arithmetic progression in $A \cap H$; since we have X_1 points and we delete at most X_2, $|B| \ge X_1 - X_2 > \frac{1}{2p}|A|$. So we can apply Ellenberg–Gijswijt now on H (viewing it as isomorphic to $\mathbb{F}_p^{n-1}$, noting that affine translations preserve arithmetic progressions), and we find that because B is a subset of H without a nontrivial 3-term arithemtic progression,

$$\frac{1}{2p}|A| < |B| \le \Gamma_p^{n-1} \implies |A| \le 2p\Gamma_p^{n-1}.$$

Since $\Gamma_p \ge 0.84p \ge \frac{2}{3}p$, this can be rewritten as $|A| \le 3(\Gamma_p)^n$, as desired. □

On our homework assignment, we'll see a problem in which we forbid configurations where a point is the center of k different arithmetic progressions, and we'll prove a similar result using a probabilistic argument.

21 April 21, 2022

We'll discuss the **Erdös-Szemerédi sunflower problem** today:

Definition 105

Three distinct sets A, B, C form a **sunflower** if $A \cap B = A \cap C = B \cap C$.

In particular, it's okay if A, B, C are all disjoint (and it's okay if one of the sets is an empty set as long as the other two are disjoint), and otherwise they form a "sunflower" (not really) shape in which the only intersection between any of the sets A, B, C is common between all three. The natural question to form from this is the following question (essentially forbidding certain patterns):

Problem 106

What is the maximum possible size of a collection of subsets of $\{1, \cdots, n\}$ (any ground set of size n works), such that $\mathcal{F}$ contains no three distinct sets A, B, C that form a sunflower?

A trivial bound is 2^n (that's the total number of subsets of $\{1, \cdots, n\}$, and the **Erdös-Szemerédi sunflower conjecture** conjectures an exponentially better bound, namely that there is some constant $c < 2$ such that we must have $|\mathcal{F}| \le c^n$. And in 2013, Alon, Shpilka, and Umans wrote a paper studying connections between problems like the sunflower problem and fast matrix multiplication (understanding how to multiply $n \times n$ matrices in time less than $O(n^3)$, motivated by the existence of certain combinatorial structures), in which they showed that **the conjecture follows** from the tri-colored sum-free theorem in $\mathbb{F}_3^n$. And after Ellenberg and Gisjwijt published their paper, it was seen that the proof there applies to the tri-colored sum-free theorem, proving the Erdös-Szemerédi sunflower conjecture. (That first appeared in print in another paper by seven different authors, discussing yet something else – the attribution of this result is generally a little complicated.)

However, it's interesting to think about whether there is a more direct way to apply the slice rank polynomial method and apply it to this problem. The answer is yes, and in fact we can improve the constant c with this direct approach:

Theorem 107 (Naslund–Sawin, 2017)

Let $\mathcal{F}$ be a collection of subsets of $\{1, \cdots, n\}$, such that no distinct subsets $A, B, C \in \mathcal{F}$ form a sunflower. Then $|\mathcal{F}| \le 3(n+1)\sum_{k \le \frac{n}{3}} \binom{n}{k}$ (which is at most 1.89^n for large enough n by a Chernoff bound calculation).

The idea is to prove a version of this problem where all subsets in $\mathcal{F}$ are of the same size – since there's only $(n+1)$ possible sizes and we're dealing with bounds that are exponential in n, this is really a lower order term. Specifically, we'll first prove the following result:

Proposition 108

Let $\mathcal{F}$ be a collection of subsets of $\{1, \cdots, n\}$ all of the same size, such that no A, B, C form a sunflower. Then $|\mathcal{F}| \le 3\sum_{k \le \frac{n}{3}} \binom{n}{k}$.

This implies the theorem, because we can consider each of the $(n+1)$ possible set sizes separately, and if we have a bound $3\sum_{k \le \frac{n}{3}} \binom{n}{k}$ on how many subsets of size s we can have, then the total number of subsets of $\{1, \cdots, n\}$ (for which $s \in \{0, \cdots, n\}$) can only be at most $(n+1)$ times this bound. (Notice that this bound is only really useful if the set size is between $\frac{n}{3}$ and $\frac{2n}{3}$ – it makes sense to ask whether we can still get an interesting bound if the set size is, for example, $\frac{n}{10}$, but there isn't an immediately clear answer.)

Proof. We'll fix some notation: let $\mathcal{F} = \{A_1, \cdots, A_m\}$, where each A_i is a subset of $\{1, \cdots n\}$. We wish to find a bound on m, and our first step is to find a polynomial on which we can apply the slice rank polynomial method. Let $x_1, \cdots, x_m$ be the indicator vectors of $A_1, \cdots, A_m$ (meaning that $x_i^{(j)} = 1$ if j is in the set A_i, and $x_i^{(j)} = 0$ otherwise). Our condition that $\mathcal{F}$ has no sunflowers is equivalent to saying that we cannot have $A_i \cap A_j = A_i \cap A_k = A_j \cap A_k$. However, notice that that intersection condition is the same as saying "any element is in either 0, 1, or 3 of the sets A_i, A_j, A_k.

In other words, for any distinct i, j, k, we require that there is some element $s \in \{1, \cdots, n\}$ that is **in exactly two** of the sets A_i, A_j, A_k, meaning that $x_i^{(s)} + x_j^{(s)} + x_k^{(s)} = 2$. Furthermore, if $i = j = k$, there will not be any s such that $x_i^{(s)} + x_j^{(s)} + x_k^{(s)} = 2$ (it's always going to be 0 or 3), and if two of i, j, k are the same (without loss of generality let's say $i = j \ne k$) then there will be some element s in $A_i = A_j$ but not in A_k (**here's where we use the fact that the sets are of the same size**), so that $x_i^{(s)} = x_j^{(s)} = 1$ but $x_k^{(s)} = 0$, so again $x_i^{(s)} + x_j^{(s)} + x_k^{(s)} = 2$. Putting this together, our sunflower-avoidance means that we can always find an s such that $x_i^{(s)} + x_j^{(s)} + x_k^{(s)}$ unless $i = j = k$, and that now looks a lot like a diagonal condition.

Motivated by this, we'll define a function $f : \{1, \cdots, m\} \times \{1, \cdots, m\} \times \{1, \cdots, m\} \to \mathbb{R}$ (any field here works as long as it's not of characteristic 2) such that

$$f(i, j, k) = \prod_{s=1}^{n} \left(x_i^{(s)} + x_j^{(s)} + x_k^{(s)} - 2 \right).$$

This is a degree n polynomial (good because it means we can use the "polynomial splitting" trick from our previous proof, and also good because the degree is not too large), and from our discussion above we can check that $f(i, j, k)$ is nonzero if and only if $i = j = k$ (that's the only case where our product has no factors of zero). So by Tao's lemma for the slice rank of a diagonal function, the slice rank must be the size of the index set, which is m.

As alluded to, we'll now get a bound on the slice rank by splitting up the polynomial into slice rank 1 functions. Since the polynomial that defines f has degree n, and each variable (meaning the components of x_i, x_j, or x_k) has degree at most 1. If we multiply out the product for $f(i, j, k)$, then for each monomial that appears, one of the $x_i^{(s)}$ variables, the $x_j^{(s)}$ variables, or the $x_k^{(s)}$ variables have sum of degrees at most $\frac{n}{3}$ by the pigeonhole principle. Now just like last time, put each monomial in the corresponding group, and then (by combining terms) we can write f as a sum of terms of the form

$$(x_i^{(1)})^{d_1} \cdots (x_i^{(n)})^{d_n} \cdot (\text{polynomial in } x_j, x_k),$$

plus similar terms $(x_j^{(1)})^{d_1} \cdots (x_j^{(n)})^{d_n} \cdot (\text{polynomial in } x_i, x_k)$ and $(x_k^{(1)})^{d_1} \cdots (x_k^{(n)})^{d_n} \cdot (\text{polynomial in } x_i, x_j)$, such that

$d_1 + \cdots + d_n \le \frac{n}{3}$ and with each $d_i \in \{0, 1\}$. Each of these terms is then a slice rank 1 polynomial (because it's a function in x_i times a function in x_j and x_k, or similar), so the slice rank must be bounded by (just counting the number of terms)

$$m = \text{slice rank}(f) \le 3 \cdot \left\{(d_1, \cdots, d_n) : d_1 + \cdots + d_n \le \frac{n}{3},\ d_i \in \{0, 1\}\right\},$$

since the slice rank is the **minimum** number of slice rank 1 polynomials required. And this is exactly $3 \sum_{k \le \frac{n}{3}} \binom{n}{k}$, since whenever $d_1 + \cdots + d_n = k$ there are $\binom{n}{k}$ ways to pick k of the d_is to be 1 and the others to be 0. □

Notice that it's important in this proof that the sum $\sum_{k \le \frac{n}{3}} \binom{n}{k}$ isn't summing over $k \le \frac{n}{2}$, for example, or else we won't get an exponentially good bound. And this detail is very similar to the Ellenberg–Gijswijt proof (in fact this proof is slightly cleaner).

Fact 109

A known lower bound for the sunflower problem, according to the Naslund–Sawin paper, is approximately 1.55^n, but this is an unpublished result by Naslund (along with lower bounds for the cap-set problem).

We'll now briefly mention another famous sunflower problem variant, **the Erdös-Rados sunflower problem**:

Problem 110

Fix some integer k. What is the maximum size of a family of sets of size k (with no restriction on the ground set), such that $\mathcal{F}$ does not contain three distinct sets A, B, C forming a sunflower?

We do need some restriction on the size of the sets – otherwise, we could use the sets $A_i = \{1, 2, \cdots, i\}$ and form an infinite set with no sunflowers. So the two natural restrictions are on the ground set and on the sizes of the sets, and we've considered both here. In this version, it's not immediately clear that this maximum size is finite, but it does turn out to be (by a combinatorial argument we won't talk about here). We'll instead say a bit about the known bounds that were established – it was conjectured by Erdös and Rado that $|\mathcal{F}| \le c^k$ for some absolute constant c, and the best known bound is recently by Alweiss, Lovett, Wu, and Zhang (2019) that $|\mathcal{F}| \le (c \log k \log\log k)^k$, greatly improving the previous bound in which the base of the exponent was polynomial in k.

Remark 111. *We've been discussing sunflowers of size* 3*, but many of these arguments can work if we have ℓ-sunflowers (sets of ℓ sets such that all pairwise intersections coincide). And if we try to take sunflowers with more sets at once, the picture looks more like an actual sunflower with petals. The problem then becomes harder, and the slice rank polynomial doesn't work anymore and no exponentially good bound is known. But it's also worth mentioning that in the Erdös-Rados sunflower problem, that proof does work more generally and just gains some factors of ℓ in the base of the exponential.*

22 May 3, 2022

In these last three lectures (as voted on by the class), we'll discuss **lower bounds for extermal numbers of bipartite graphs** (via randomized and algebraic constructions). To get in the mood for these kinds of questions, we'll start with a classic problem from more than a century ago:

Problem 112

What is the maximum number of edges that an n-vertex graph can have without containing a triangle?

This is a classic problem that we may have seen before – the answer is $\lfloor \frac{n^2}{4} \rfloor$, obtained from a bipartite graph with $\lfloor \frac{n}{2} \rfloor$ vertices on one side and $\lceil \frac{n}{2} \rceil$ on the other. We won't go over the proof here, but it's something we can search up – instead, we'll generalize the problem:

Definition 113

Let H be a fixed graph. For any n, let $\text{ex}(n, H)$ be the maximum number of edges of an n-vertex graph without a copy of H as a subgraph.

(For example, if H is the triangle graph, then $\text{ex}(n, \Delta) = \lfloor \frac{n^2}{4} \rfloor$. And it's important that we're asking H to not be contained as a subgraph, rather than as an induced subgraph – after all, if H is not a clique, we could just take the complete graph on n vertices and we would not have a copy of H as an induced graph.)

Our first result gives us an asymptotic fraction of edges that we may include as n gets large:

Theorem 114 (Erdös–Stone–Simonovits)

For any fixed graph H, let $\chi(H)$ denote the chromatic number of H (the minimum number of colors needed to color the vertices of H such that any two adjacent vertices have different colors). Then

$$\text{ex}(n, H) = \left(1 - \frac{1}{\chi(H) - 1}\right) \binom{n}{2} + o(n^2).$$

To get some intuition for this result, the idea is to split our n vertices into $\chi(H) - 1$ equal-sized parts, and make our graph empty within each part and complete between the parts (so this is a "blow-up" of the clique on $\chi(H) - 1$ vertices). That graph can be colored with $\chi(H) - 1$ colors, so it cannot contain H (which requires $\chi(H)$ colors to properly color). We won't prove this result because it's not directly related to our goal, though.

This result may look like it gives us a good sense of the number of allowed edges, but in the case where H is bipartite (and thus the chromatic number is 2), we don't actually get an asymptotic behavior – in that case we just know that $\text{ex}(n, H) = o(n^2)$ and the problem is in fact still open. In fact, even for the most simple bipartite graphs, the complete bipartite graphs $K_{s,t}$ (in which we have s vertices on one side, t vertices on the other, and the only edges are the st edges between the two sides), the problem is very difficult and still mostly open.

We'll discuss this case $H = K_{s,t}$ in class today, though, starting with an upper bound:

Theorem 115 (Kővári–Sós–Turán (1954))

Fix $s \le t$. Then there exists a constant $C_{s,t}$ such that for all n, we have

$$\text{ex}(n, K_{s,t}) \le C_{s,t} n^{2-1/s}.$$

In particular, this answer is indeed always $o(n^2)$ (as predicted by Theorem 114), and we get a weaker bound for larger s (which makes sense, because larger s means we have a larger subgraph to avoid, which is easier to do).

Proof. To show this upper bound, we must show that any graph G on n vertices that does not contain $K_{s,t}$ has at most $C_{s,t} n^{2-1/s}$ edges. We'll count the number of $(s+1)$-tuples $(x_1, \cdots, x_s, y) \in V(G)^{s+1}$, such that each x_i is connected by an edge to y (in other words, (x_i, y) is an edge). Call this number N. We know that

$$N = \sum_{(x_1, \cdots, x_s) \in V(G)^s} |\{y \in V(G) : (x_i, y) \in E(G) \quad \forall 1 \le i \le n\},$$

and now we can break this up into two cases: if the x_is are all distinct, then we cannot have more than $(t-1)$ options for y, or else we'd contain a copy of $K_{s,t}$, and in general we can only have n vertices in the graph, so this is

$$\leq \sum_{\substack{(x_1,\cdots,x_s)\in V(G)^s \\ \text{distinct}}} (t-1) + \sum_{\substack{(x_1,\cdots,x_s)\in V(G)^s \\ \text{not distinct}}} n \leq n^s \cdot t + \binom{s}{2} n^{s-1} \cdot n \leq (t+s^2)n^s$$

(crudely bounding both terms, with the latter bound $\binom{s}{2}n^{s-1}$ coming from choosing two indices to be equal and doing a union over all possible choices). On the other hand, we may count N by starting off with y (at which point each of the x_is can be any of $\deg(y)$ choices):

$$N = \sum_{y\in V(G)} (\deg y))^s = n \cdot \frac{1}{n} \sum_{y\in V(G)} (\deg(y))^s,$$

and now we make use of the convexity of the function $z \mapsto z^s$ through Jensen's inequality (this is also the power mean inequality):

$$\geq n \cdot \left(\frac{1}{n} \sum_{y\in V(G)} \deg(y) \right)^s = n^{1-s} \left(\sum_{y\in V(G)} \deg(Y) \right)^s = n^{1-s}(2|E(G)|)^s.$$

Putting our inequalities together, we find that

$$2^s n^{1-s}|E(G)|^s \leq N \leq (t+s^2)n^s \implies |E(G)|^s \leq 2^s(t+s^2)n^{2s-1},$$

and taking sth powers gives us the desired result (with $C_{s,t} = 2(t+s^2)^{1/s}$). □

Fact 116

It's conjectured that this result is tight up to the constant factor (because this proof strategy is really the only one that's known). That's known for $K_{2,2}$, which implies the result for $K_{2,t}$ (because if we can construct an example for $K_{2,2}$, that example also works for $K_{2,t}$ and we still have the same bound in the theorem), as well as for $K_{3,3}$ and thus $K_{3,t}$ (Brown, 1966). But the true upper bound for $K_{4,4}$ is unknown, and the only other thing known (Bukh, 2021) is that if $t > 9^s s^{4s^{2/3}}$ (which is basically $9^{s+o(s)}$) we also have a tight upper bound (though 20 years ago it was known for $t \geq s! + 1$ and then $t \geq (s-1)! + 1$).

All of the constructions above are algebraic, but some of them make use of a "randomized algebraic method." To motivate that a bit and understand how we construct these upper bounds, consider the following setting:

Example 117

Let p be a large prime. We will construct a bipartite graph G, such that the left and right vertex set are both $\mathbb{F}_p^s$, and such that G does not contain any copies of $K_{s,t}$.

We may have a few concerns with this setup, which we'll address now. First of all, notice that we only lose at most a factor of 2 when we go from a general graph G to a bipartite one, because we can always take a random bipartition of G, meaning that each vertex of G goes into one of the two sides and we delete all edges within that set. We keep each edge with probability $\frac{1}{2}$ through this, so there is indeed always a bipartition for which at least half the edges are kept. Second, even though this construction only allows us graphs with $n = 2p^s$ vertices (where p is a prime), notice that for any $n \in \mathbb{N}$, there is some p such that $\frac{n}{2^{s+1}} \leq p^s \leq \frac{n}{2}$ (by Bertrand's postulate, there is always a prime between m and $2m$). So we can basically take $2p^s$ of the vertices and do this construction, isolate the remaining $n - 2p^s$

vertices, and because n is only off from $2p^s$ by at most a constant factor 2^s we don't lose any asymptotic strength here either.

The construction works as follows (and this is where the algebraic structure comes in): for any $(x, y) \in \mathbb{F}_p^s \times \mathbb{F}_p^s$, we decide whether to include the edge between x and y via an algebraic construction. This is a bit complicated, so we'll describe some simple cases for illustration:

- For $K_{2,2}$, we have $x = (x_1, x_2)$ and $y = (y_1, y_2)$, and we include an edge between x and y if and only if $x_1y_1 + x_2y_2 = 1$.
- For $K_{3,3}$ and $p \equiv 3 \bmod 4$, we connect (x_1, x_2, x_3) to (y_1, y_2, y_3) if and only if $(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2 = 1$.
- For $K_{s,t}$ with $t \geq s! + 1$ (one of the earlier bounds in Fact 116), instead of taking $\mathbb{F}_p^s$ we in fact want to take $\mathbb{F}_{p^s}$ (which has additive structure isomorphic to $\mathbb{F}_p^s$ but also has a field structure). Then we include an edge $(x, y) \in \mathbb{F}_{p^s} \times \mathbb{F}_{p^s}$ if and only if $\text{Norm}(x + y) = 1$ in $\mathbb{F}_p$. (If we haven't taken a Galois theory class, the norm of an element $\alpha \in \mathbb{F}_{p^s}$ is the product of all Galois conjugates of α. But alternatively, $\text{Norm}(\alpha)$ is the determinant of the $\mathbb{F}_p$-linear map $\mathbb{F}_{p^s} \to \mathbb{F}_{p^s}$ which sends z to αz, and that explains why $\text{Norm}(\alpha)$ is always in $\mathbb{F}_p$.)

It's difficult to analyze the latter two constructions, but we can do so for $K_{2,2}$ and we will do so now (remember we're doing this to check that Kővári–Sós–Turán is tight for $s = 2$).

- First, we calculate how many vertices and edges our construction gives. We have $n = 2p^2$ vertices, and our goal is to get $O(p^3)$ edges (since $2 - 1/s = 3/2$ and we want $n^{3/2}$ edges up to a constant factor). If $(x_1, x_2) = (0, 0)$, then there are no edges from x to any of the other vertices. But otherwise, we always have p choices for (y_1, y_2) because we have a nontrivial linear equation in $\mathbb{F}_p^2$, cutting out a line. Thus, we indeed have $p(p^2 - 1) \approx p^3$ edges in this graph we've constructed.
- Now we must check that there is no $K_{2,2}$ in this graph. In other words, we must show that for any distinct (x_1, x_2) and (x_1', x_2'), there is at most one (y_1, y_2) satisfying $x_1y_1 + x_2y_2 = 1$ and $x_1'y_1 + x_2'y_2 = 1$. (Normally we would need to make sure that there aren't any other edges in the induced subgraph H, but because we constructed G to be bipartite and $K_{2,2}$ is complete bipartite we don't need to worry about that.) But these are either linearly independent equations or (x_1, x_2) is a multiple of (x_2', x_2'), so there must be only one or zero solutions, respectively.

Remark 118. *We can think about this $K_{2,2}$ construction by instead thinking about the* **finite projective plane** *over $\mathbb{F}_p$, where the vertices on the left are the points in the projective plane, the vertices on the right are the lines in the projective plane, and edges are drawn corresponding to incidences. This is basically the previous construction "deleting the zeros," and it's a nice tool for generating constructions in various situations. (In fact, this is how the game Dobble generates its cards, each containing 6 or 8 symbols, so that each pairs of cards has exactly one symbol in common.)*

Remark 119. *The reason we cannot use $x_1y_1 + x_2y_2 + x_3y_3 = 1$ for the $K_{3,3}$ case is that (even though we get the right number of edges) we will not have a $K_{3,3}$-free graph is that we can pick three (x_1, x_2, x_3) triples so that the corresponding hyperplanes all intersect along a common line. Then picking any three points (y_1, y_2, y_3) along that line gives us a $K_{3,3}$. Instead, the $(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 = 1$ construction basically comes down (non-rigorously) to the fact that three unit spheres can only intersect in two points.*

This kind of construction is hard to do for general s because it requires us to construct specific polynomials to establish the algebraic condition, and the reason we have a new bound $t > 9^s s^{4s^{2/3}}$ is because there is a nice way to do so **randomly**. (This type of approach started becoming popular recently – for example, a randomized algebraic construction was used to prove a major result in combinatorial design theory.) What we'll see next time is how such a construction works for our avoiding-bipartite-graphs problem!

23 May 5, 2022

Remark 120. *As a reminder, we should all fill out the subject evaluations (this is sort of like a "grade for the instructor").*

Last lecture, we studied the quantity $\text{ex}(n, H)$ (the maximum number of edges in an n-vertex graph without a copy of H as a subgraph) where H is the complete bipartite graph $K_{s,t}$. (Erdös-Stone-Simonovits already tells us the asymptotic behavior for the case where H is not bipartite.) We proved the upper bound $\text{ex}(n, K_{s,t}) \le C_{s,t}n^{2-1/s}$ for any $s \le t$, and today we'll find a matching lower bound for sufficiently large t (relative to s). (Remember that it's easier to avoid $K_{s,t}$ if t is larger, so ideally we would do this for $K_{s,s}$ but that's only known for $s = 2, 3$.) We will follow Bukh's argument from 2014; although Kollár, Rónyai, and Szábo (1996), and Alon, Rónyai, and Szábo (1999) already previously yielded better results, those bounds were obtained through explicit constructions that do not generalize. In contrast, Bukh's argument is more general and was in fact recently improved (in 2021) to $t > 9^s(s^4s^{2/3})$, better than the previously known $t > (s-1)! + 1$.

Theorem 121

If t is sufficiently large with respect to s, then $\text{ex}(K_{s,t}) \ge c_{s,t}n^{2-1/s}$.

Proof. To prove this result, we must construct an example of a graph on n vertices with $c_{s,t}n^{2-1/s}$ edges not containing any copies of $K_{s,t}$. Like last lecture, we may assume $n = 2p^s$ for some prime p by Bertrand's postulate, and we may construct our graph to be bipartite – these will only contribute constant-in-n factors to the bound. Rewriting the goal in terms of p, we wish to construct a bipartite graph G with p^s vertices on each side, with at most $c_{s,t}p^{2s-1}$ edges (possibly a different $c_{s,t}$ than above), and with no copies of $K_{s,t}$.

To do this, we identify each side of our graph with $\mathbb{F}_p^s$ (s-tuples of $\mathbb{F}_p$), and we connect x on the left side with y on the right side with some algebraic condition. We did this last time for the $K_{2,2}$ case with an explicit example of the polynomial, but Bukh's key insight is that we can now pick a **random polynomial** $f \in \mathbb{F}_p[x_1, \cdots, x_s, y_1, \cdots, y_s]$ and connect x and y in G if and only if $f(x_1, \cdots, x_s, y_1, \cdots, y_s) = 0$. However, we'll need to be a bit more precise with this – there are only finitely many possible evaluation functions, but we're going to want to think of f as an actual polynomial rather than just as a function.

Specifically, we'll let $d = s^3 - 1$ (we'll see why we choose this particular value later), and among all polynomials in $\mathbb{F}_p[x_1, \cdots, x_s, y_1, \cdots, y_s]$ of degree at most d, we pick one uniformly at random. (There are various variations on how to set this up exactly – we could use homogeneous polynomials, or require that the degrees in x and y are the same – but this is a relatively simple one to state.) Naively, for any x and y it makes sense that $f(x, y) = 0$ occurs with probability $\frac{1}{p}$, but we need to be more precise with that if we want to consider graphs of the form $K_{s,t}$:

Lemma 122

Let $(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)}) \in \mathbb{F}_p^s \times \mathbb{F}_p^s$ be distinct pairs of points with $m \le d + 1$ (though the same $x^{(i)}$ can show up multiple times in different pairs). Then

$$\mathbb{P}\left(f(x^{(i)}, y^{(i)}) = 0 \quad \forall 1 \le i \le m\right) = \frac{1}{p^m}.$$

In other words, if we have sufficiently few points, there won't be nasty dependencies in the probabilities of vanishing. (And we can really just think of each $(x^{(i)}, y^{(i)})$ as a point in $\mathbb{F}_p^{2s}$ and evaluating at a polynomial f in $2s$ variables.)

Proof of lemma. Notice that for each $1 \le i \le m$, we can find a polynomial $P_i \in \mathbb{F}_p[x_1, \cdots, x_s, y_1, \cdots, y_s]$ of degree at most $m - 1 \le d$, such that $P_i(x^{(j)}, y^{(j)}) = 1$ if $i = j$ and 0 otherwise. (This is because for each $j \ne i$, we may pick a hyperplane (degree-1 polynomial) through $(x^{(j)}, y^{(j)})$ but not through $(x^{(i)}, y^{(i)})$, and multiply all $(m-1)$ such hyperplanes together to get a polynomial vanishing on all points except $x^{(i)}, y^{(i)}$, then rescale.) The polynomials $P_1, \cdots, P_m$ are then linearly independent, so we can extend $\{P_1, \cdots, P_m\}$ to a basis of all polynomials f of degree at most d, which we'll write $\{P_1, \cdots, P_k\}$ (where $k = \binom{2s+d}{d}$).

We can now pick f uniformly at random by picking uniformly random coefficients for each element of the basis – while a priori it might be more natural to pick coefficients for the monomials at random, for this problem it's more insightful to pick $a_1, \cdots, a_k \in \mathbb{F}_p$ independently and uniformly at random and set $f = a_1P_1 + \cdots + a_kP_k$. And now if we pick $a_{m+1}, \cdots, a_k$ first, the polynomial $g = a_{m+1}P_{m+1} + \cdots + a_kP_k$ takes some values on the evaluation points $(x^{(1)}, y^{(1)}), \cdots, (x^{(m)}, y^{(m)})$. But no matter whatever coefficients we choose, the probability that $f(x^{(i)}, y^{(i)}) = 0$ now only depends on a_i (not any of the other coefficients $a_1, \cdots, a_m$) and occurs with probability $\frac{1}{p}$. Since the a_is are independent, this gives the desired result. □

We can now return to the properties of G. The expected number of edges of G when we pick f randomly (applying Lemma 122 with $m = 1$ and using linearity of expectation) is $p^s \cdot p^s \cdot \frac{1}{p} = p^{2s-1}$. The idea is that we wish to show that G will have few copies of $K_{s,t}$ (in particular, at most half as many as p^{2s-1}), because then we can get a new graph with no copies of $K_{s,t}$ by **deleting an edge from each $K_{s,t}$ subgraph**.

To control the number of copies of $K_{s,t}$, we can just look at those with s vertices on the left and t vertices on the right (we will account for a factor of 2 later to count the other way around by symmetry). Having a $K_{s,t}$ means that if we look at a set of s vertices on the left and look at their neighborhoods, those neighborhoods intersect in at least t points. (So we won't exactly count the number of $K_{s,t}$s, only count the number of problematic s-sets, and we'll delete all the neighbors of a given left vertex to fix those problematic sets.) For any subset U on the left of size s, we define $N(U) \subseteq \mathbb{F}_p^s$ to be the set of its common neighbors. For any fixed U, we then have

$$\mathbb{E}[|N(U)|] = p^s \cdot \frac{1}{p^s} = 1,$$

because for any point on the right we can apply Lemma 122 to find that there is a probability $\frac{1}{p^s}$ that it's connected to all of U (and then we use linearity of expectation). (Here it's important that $s \le d+1$.) We can then say that by Markov's inequality, only a fraction $\frac{1}{t}$ of the sets U can be bad, but that's not good enough for us because it's still a constant fraction of all possible sets (and we should remember that p is very large relative to t, even if t is very large relative to s). So we need to "enhance" Markov's inequality by looking at the qth moment, and it will turn out that $q = s^2$ is the right choice. Notice that $|N(U)|^q$ is the number of q-tuples of points, not necessarily distinct, on the right that are all connected to elements of U, so

$$\mathbb{E}[|N(U)|^q] = \mathbb{E}\left[(y_1, \cdots, y_q) \in (\mathbb{F}_p^s)^q : y_1, \cdots, y_q \text{ common neighbors of } U\right].$$

We now claim that this expectation is at most a constant depending on s. (This is a bit trickier because the number of edges in the constraint may vary depending on whether there are repetitions in the y_is.) For any fixed q-tuple $(y_1, \cdots, y_q) \in (\mathbb{F}_p^s)^q$ containing k different points (for some $1 \le k \le q = s^2$), the probability that $y_1, \cdots, y_q$ are all common neighbors of U is $\frac{1}{p^{sk}}$ (because there are sk different edges that must all be present for this to occur) by Lemma 122, since $sk \le d + 1 = s^3$. Since the number of q-tuples $(y_1, \cdots, y_q) \in (\mathbb{F}_p^s)^q$ with at most k different entries is $\text{const}(k, q)p^{sk}$, which is a constant depending on s times p^{sk}, we find that

$$\mathbb{E}[|N(U)|^q] = \sum_{k=1}^{q} \text{const}(s)p^{sk} \cdot \frac{1}{p^{sk}} \le C_s$$

for some constant C_s, as desired. This is still not good enough – applying Markov's still gives us a constant fraction

$\frac{C_s}{t^q}$ of the total possible sets U that can be problematic, and this still has no p-dependence. But **here's where the algebra comes in** – for every fixed x, $f(x_1, \cdots, x_s, y_1, \cdots, y_s) = 0$ is a polynomial condition in y, so we are cutting out an **algebraic set** (in other words, a **variety**) when we require a point to be in $N(U)$. And intuitively, the idea is that varieties have dimensions which constrain the number of points that can be in the set $N(U)$:

Fact 123

Suppose $f_1, \cdots, f_k \in \mathbb{F}_p[y_1, \cdots, y_s]$ are polynomials of degree at most d (not necessarily the same d as above). Then the set $W = \{y \in \mathbb{F}_p^s : f_1(y) = \cdots = f_k(y) = 0\}$ either has $|W| \le \text{const}(s, k, d)$ or $|W| \ge p - \text{const}(s, k, d)\sqrt{p}$ (in particular, $|W| \ge \frac{p}{2}$ for sufficiently large p).

(This is essentially the Lang-Weil bound – usually this kind of result only holds for irreducible varieties over algebraically closed fields, but there are ways to generalize it.) Now for a fixed $U = \{x^{(1)}, \cdots, x^{(s)}\}$, we indeed have $N(U) = \{y \in \mathbb{F}_p^s : f(x^{(1)}, y) = f(x^{(2)}, y) = \cdots = f(x^{(s)}, y) = 0\}$ (so in other words, we define $f_i(\cdot) = f(x^{(i)}, \cdot)$). So either $|N(U)| \le C_s$ or $|N(U)| \ge \frac{p}{2}$, and now we can make use of our Markov bound. Assuming that $t > C_s$ (the constant from our q-moment bound above), we now have

$$\mathbb{P}(|N(U)| > t) = \mathbb{P}\left(|N(U)| \ge \frac{p}{2}\right) = \mathbb{P}\left(|N(U)|^q \ge \left(\frac{p}{2}\right)^q\right) \le \frac{\mathbb{E}[|N(U)|^q]}{(p/2)^q} \le C_s' p^{-s^2}.$$

by Markov's inequality. Since this held for a fixed U, we find that the **expected** number of sets U with $|N(U)| \ge t$ is at most $(p^s)^s \cdot C_s' p^{-s^2} = C_s'$. Similarly, the expected number of sets U on the right side of size s with $|N(U)| \ge t$ is also at most that constant. So if we now make this "deleting" argument, where we delete all edges adjacent to a particular bad vertex for each U,

$$\mathbb{E}[|E(G)| - p^s \cdot \text{number of bad} U] \ge c_{s,t} p^{2s-1} - 2C_s' p^s \ge \frac{c_{s,t}}{2} p^{2s-1}$$

for sufficiently large p. Thus we can pick some function f such that deleting these bad edges gives us a graph with at least $\frac{c_{s,t}}{2} p^{2s-1}$ edges and no copies of $K_{s,t}$, as desired. □

24 May 10, 2022

We'll continue last week's discussion on extremal numbers of bipartite graphs today. In particular, we've previously discussed the Kövari-Sós-Turán upper bound on $\text{ex}(n, K_{s,t})$ (with a double-counting argument), and we've also obtained a matching lower bound (up to constants) on $\text{ex}(n, K_{s,t})$ (with a probabilistic argument).

Fact 124

To discuss what's known in more detail for general bipartite graphs, the value of $\text{ex}(n, H)$ is known up to constant factors for $K_{s,t}$ (as discussed in lecture) for $s = 2, 3$ or $t > 9^s s^{4s^{2/3}}$ (though we did not show this strongest bound). Other than that, we also know $\text{ex}(n, H)$ for (1) a tree, (2) a cycle of length 4, 6, or 10, or (3) a collection of ℓ paths of length k all connected at the start and end point for sufficiently large ℓ relative to k. (In particular, this is a single cycle if $\ell = 2$.) The answers are (1) $\Theta_H(n)$ (meaning linear in n with constant factor depending on H), (2) $\Theta_k(n^{1+1/k})$, and (3) $\Theta_{k+\ell}(n^{1+1/k})$, respectively.

Recall that for graphs H that are not bipartite, $\text{ex}(n, H)$ is asymptotically proportional to n^2, and for all of the bipartite graphs above we have $\text{ex}(n, H) = \Theta_H(n^\alpha)$ for some rational number $1 \le \alpha \le 2$. We can check that for any graph H with more than one edge, we must have $1 \le \alpha \le 2$ if the answer is of this form, but trying to prove that α is indeed rational or that the asymptotic behavior is still open. And there is in fact a relevant "backwards" conjecture here (still widely open, though there is active progress being made):

Conjecture 125 (Erdös-Simonovits)

For every rational $1 \le \alpha \le 2$, there is some graph H with $\text{ex}(n, H) = \Theta_H(n^\alpha)$.

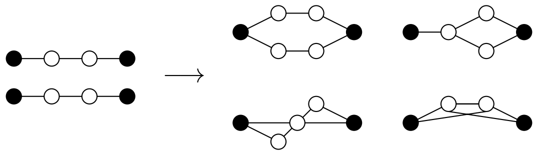
However, we can also extend our definition $\text{ex}(n, H)$ to **finite families** $\mathcal{H}$ of graphs and let $\text{ex}(n, \mathcal{H})$ be the maximum number of edges in an n-vertex graph which contains no graph $H \in \mathcal{H}$. Then there are results that are known:

Theorem 126 (Bukh–Conlon (2018))

For every rational $1 \le \alpha \le 2$, there is a finite family $\mathcal{H}$ of graphs such that $\text{ex}(n, \mathcal{H}) = \Theta_{\mathcal{H}}(n^\alpha)$.

(For $\alpha = 1$ we can take a tree, and for $\alpha = 2$ we can take any non-bipartite graph, so we only need to consider rational numbers strictly between 1 and 2.) In order to understand this result, we'll describe a construction. Consider a tree T, and let $R \subseteq V(T)$ be a subset of the vertices with no edges between vertices in R (in the paper, these are called the "roots" of the tree, because we can imagine placing the vertices in R at the bottom – all roots will turn out to always be leaves in our construction). Normally we would call these "independent sets," but we will reserve the use of "independent" for probabilistic arguments later.

We'll then consider many copies of the same tree and **glue along roots** – specifically, we'll consider all ways that we can glue t different copies of T along the set R, such that all t copies of any $r \in R$ are identified, and different non-root vertices may be identified as well as long as we don't literally glue two copies of T on top of each other. Here's a diagram of some ways in which we may identify two copies of a path of length 4:



Let $\mathcal{H}^{(t)}_{(T,R)}$ be the family of graphs that may be obtained in this way.

Definition 127

For a tree with roots (T, R), let the **density** $\rho(T, R)$ be

$$\rho(T,R) = \frac{|E(T)|}{|V(T)| - |R|}.$$

The idea with this quantity is that $|V(T) - |R|$ is the number of "additional vertices" beyond the roots that each new copy of T may have, and this density is always at least 1.

Definition 128

A tree (T, R) is **balanced** if for every subset $S \subseteq V(T) \setminus R$, we have

$$\frac{\text{number of edges with at least one vertex in } S}{|S|} \ge \rho(T,R).$$

In other words, there are always a reasonable number of edges touching S – for example, taking $S = V(T) \setminus R$ itself, we get equality because all edges touch at least one vertex in $V(T) \setminus R$. And we're now ready to state the main result we'll prove today:

Theorem 129 (Bukh–Conlon)
Suppose (T, R) is a balanced tree. Then for sufficiently large $t \in \mathbb{N}$, we have

$$\text{ex}\left(n, \mathcal{H}_{(T,R)}^{(t)}\right) = \Theta_{T,R,t}\left(n^{2-\frac{1}{\rho(T,R)}}\right).$$

In particular, if we can show this result, then we can prove Theorem 126 by showing that we can have $\rho(T, R)$ be any rational number larger than 1 (by attaching roots to a tree in a way that vaguely resembles the Euclidean algorithm but is completely explicit – we won't do it here).

Example 130
Suppose (T, R) is a star graph with s leaves and a single center vertex, and set R to be the set of all leaves. Then notice that $\mathcal{H}_{(T,R)}^{(t)} = \{K_{s,t}\}$ (each center vertex must be different when we overlay the graphs), and if we calculate $\rho(T, R)$ we will find that this is in agreement with the results we proved last week.

Example 131
Suppose (T, R) is a path of k edges with R being the two endpoints of the path. We can check that this graph is balanced, which gives us a generalization of part (3) of Fact 124. Specifically, Bukh and Conlon's result gives us the **lower bound** because it constructs an explicit example of a graph which ignores the ℓ paths of length k all glued together at the endpoints.

We'll just do the proof of the lower bound – the upper bound essentially comes from considering the average degree of the vertices and constructing a subgraph in which all vertices have at least half that degree (only losing a constant factor), at which point we can construct many copies of T and thus some choice of root vertices R will give us a graph in $\mathcal{H}_{(T,R)}^{(t)}$.

This proof will follow a similar randomized algebraic construction as last lecture's proof, and we need some technical results first:

Lemma 132
For any balanced tree (T, R), any $H \in \mathcal{H}_{(T,R)}^{(t)}$ has "many vertices:"

$$|E(H)| \geq \rho(T, R)(|V(H)| - |R|).$$

(For example, if we glue all copies of T distinctly, we get equality.) Essentially, the proof of this is induction on the size of t: the base case $t = 1$ is clear, and for the inductive hypothesis we're adding a new copy of T. But if S is the set of new vertices, we introduce $|S|$ new vertices and at least $\rho(T, R)|S|$ new edges because the tree is balanced.

Our next result is similar to the result about sizes of varieties that we used at the end of last class, and it follows from algebraic geometry (again this is because of having a "well-defined dimension" in some sense):

Lemma 133

Suppose $f_1, \cdots, f_k, g_1, \cdots, g_\ell \in \mathbb{F}_p[y_1, \cdots, y_s]$ are polynomials of degree at most d, and define

$$W = \{y \in \mathbb{F}_p^s : f_1(y) = \cdots = f_k(y) = 0\}, \quad D = \{y \in \mathbb{F}_p^s : g_1(y) = \cdots = g_\ell(y) = 0\}.$$

Then the size of the set $W \setminus D$ satisfies either $|W \setminus D| \le \text{const}(s, k, \ell, d)$ or $|W \setminus D| \ge p - \text{const}(s, k, \ell, d)\sqrt{p}$ (in particular, in this latter case we have $|W \setminus D| \ge \frac{p}{2}$ for sufficiently large p).

(The idea is to reduce to the "irreducible" case, in which we know that the intersection of W and D is either W itself or of a lower dimension.)

Proof of the lower bound of Theorem 129. To make our notation easier, we'll define $a = |V(T)| - |R|$ and $b = |E(T)|$, so that $\rho(T, R) = \frac{b}{a}$, and we'll also define $r = |R|$. Label the vertices of T by the set $\{1, \cdots, a + r\}$, such that the first r vertices are the roots in R.

Let $d = bq$ and $q = 2br$ (if we remember from last time, d will end up being the degree of our polynomial, and q will be the moment that we consider). We need to construct a graph G of n vertices and $\Omega(n^{2-a/b}) = \Omega(n^{(2b-a)/b})$ edges, and we'll do this, like last time, by constructing a bipartite graph G of $2p^b$ vertices identified with two copies of $\mathbb{F}_p^b$ for some prime power p – remember that doing this only loses us a constant factor. And this time, we'll decide whether we have edges between vertices by looking at multiple random polynomials (instead of just one): letting $f_1, \cdots, f_a \in \mathbb{F}_p[x_1, \cdots, x_b, y_1, \cdots, y_b]$ be independent random polynomials of degree at most d, we will draw an edge between x and y if and only if $f_1(x, y) = \cdots = f_a(x, y) = 0$.

Since $f_i(x, y) = 0$ occurs with probability $\frac{1}{p}$, the expected number of edges in our graph is

$$\mathbb{E}[|E(G)|] = (p^b)^2 \cdot \left(\frac{1}{p}\right)^a = p^{2b-a}.$$

(Notice that this is precisely the order of the number of edges that we want, since $n = 2p^b$.) We now need to avoid elements of $\mathcal{H}^{(t)}_{(T,R)}$, and we're going to do this by **counting the number of potentially bad r-tuples** that could potentially be roots. In other words, mirroring the proof from last time, fix any $(z_1, \cdots, z_r) \in \mathbb{F}_p^b$ (as a set of potential roots) and fix z_1 to be on the left side of the graph (we'll need to put in an additional factor of 2 later). Since T is a tree, this fixes which side each root z_i will be on, and now define $M(z_1, \cdots, z_r)$ to be the number of copies of T appearing in G, with root 1 on the left and with root i mapped to z_i for all $1 \le i \le r$ (this is a random variable). Our ultimate goal is to avoid having $|M(z_1, \cdots, z_r)| > t$, because that would give us t copies of T with the same roots R and thus give us an element of $\mathcal{H}^{(t)}_{(T,R)}$. We have

$$\mathbb{E}[M(z_1, \cdots, z_r)] \le (p^b)^a \cdot (p^{-a})^b = 1,$$

because intuitively there are at most (p^b) potential places where each non-root vertex can go (which side it's on is fixed by the choice of root 1), there are a such non-root vertices, and each edge occurs with probability p^{-a}. (Actually, because of independence arguments, we needed to use the result from last lecture, which says that the probability any of the a polynomials vanishes on all b edges is p^{-b} because b is sufficiently small.) And now just like last time, we do a moment bound: thinking about q-tuples of these potential trees, we sum over how many (q') distinct different trees actually occur, and we find that (after some calculation mirroring last lecture)

$$\mathbb{E}[M(z_1, \cdots, z_r)^q] \le \sum_{q' \le q} (q')^q \sum_{H \in \mathcal{H}^{(q')}_{(T,R)}} (p^b)^{|V(H)| - |R|} \cdot \left(p^{-a}\right)^{|E(H)|}.$$

But now we can use the lower bound from Lemma 132 and find that this expectation is independent of n and is a constant $\text{const}(T, R)$. And like last time, Markov's inequality alone is not enough, but we will apply Lemma 133 here – copies of T that contribute to M are indeed solutions of a combination of polynomial relations in $z_{r+1}, \cdots, z_{r+a}$, because for example requiring that 1 and $r+1$ are connected in the tree requires $f_1(z_1, z_{r+1}) = \cdots = f_a(z, z_{r+1}) = 0$, and for example requiring that $r+1$ and $r+2$ be not connected requires $f_1(z_{r+1}, z_{r+2}) = \cdots = f_a(z_{r+1}, z_{r+2}) \neq 0$ (or, if $r+1$ is on the right and $r+2$ is on the left in our fixing of the tree, $f_1(z_{r+2}, z_{r+1}) = \cdots = f_a(z_{r+2}, z_{r+1}) \neq 0$). Additionally, to make sure we actually count the exact quantity $M(z_1, \cdots, z_r)$ is counting, we have to also make sure that we don't have a "degenerate" situation where we use the same vertex twice. So in particular, we must also require that whenever i and j are fixed to be on the same side of the tree for $r+1 \leq i \neq j \leq r+a$, we have $z_i - z_j \neq 0$.

With that, the set of points in $M(z_1, \cdots, z_r)$ is indeed restricted as in Lemma 133, so we then find that $M(z_1, \cdots, z_r)$ is either at most $\text{const}(T, R)$ or at least $\frac{p}{2}$, so we can choose $t > \text{const}(T, R)$. And just like last time, we therefore get by Markov's inequality that

$$\mathbb{P}(M(z_1, \cdots, z_r) \geq t) = \mathbb{P}\left(M(z_1, \cdots, z_r) \geq \frac{p}{2}\right) \leq \frac{\mathbb{E}[(M(z_1, \cdots, z_r)^q]}{(p/2)^q} \leq \text{const} \cdot p^{-q} = \text{const} p^{-2br}.$$

Thus (remembering to count both left and right sides) the expected number of events where $M(z_1, \cdots, z_r) \geq t$ over all possible $(z_1, \cdots, z_r)$s (and also accounting for z_1 being potentially on the left or the right) is at most $2(p^b)^r \cdot \text{const} p^{-2br}$, which is in particular bounded by a constant. The rest of the argument is now identical to last time: we have p^{2b-a} edges in G in expectation, and if we delete all edges from one of the vertices in each $(z_1, \cdots, z_r)$ with $M(z_1, \cdots, z_r)$ too large, we delete at most cp^a edges in expectation. Since $a < b$, the number of deleted edges is small for large p, and thus (for sufficiently large t and p) there is some choice of polynomials $f_1, \cdots, f_a$ such that we have $\Theta(p^{2b-a}) = \Theta(n^{(2b-a)/b}) = \Theta(n^{2-\frac{1}{\rho(T,R)}})$ edges and avoid all subgraphs of $\mathcal{H}^{(t)}_{(T,R)}$, as desired. □