18.225 — Graph Theory and Additive Combinatorics

Class by Yufei Zhao

Notes by Sanjana Das

Fall 2023

Notes for the MIT class **18.225** (Graph Theory and Additive Combinatorics), taught by Yufei Zhao. All errors are my responsibility.

Contents

§1 September 6, 2023

A hundred years ago, many mathematicians were interested in solving Fermat's last theorem. One of the approaches at the time was to look at Fermat's equation — $X^n + Y^n = Z^n$ — mod p. If you can show this equation has no solutions mod p, then it has no nontrivial integer solutions. Unfortunately (or fortunately), this method doesn't work; we'll start the class by explaining a proof of why this approach is doomed to fail (which will lead to the theme of the class — a connection between graph theory and additive combinatorics).

Theorem 1.1 (Dickson 1909)

For every n, the equation $X^n + Y^n \equiv Z^n \pmod p$ has nontrivial solutions for all sufficiently large primes p.

Schur later gave a proof of the same fact using a much simpler proof, demonstrating an important combinatorial technique.

Theorem 1.2 (Schur's theorem)

If $\mathbb{N}$ is colored using finitely many colors, then there exists a monochromatic solution to the equation $x + y = z$.

§1.1 Finitary vs. Infinitary

Today we'll do two things — explain how to prove Schur's theorem, and explain why it implies the earlier statement about FLT mod p.

Before that, as with many statements we'll see in this class, statements like this have a different but equivalent formulation, referred to as the *finitary* version.

Theorem 1.3 (Schur's theorem, finitary version)

For every r, there exists N (depending on r) such that if $[N] = \{1, \dots, N\}$ are colored using r colors, then there exists a monochromatic solution to the equation $x + y = z$.

These look similar, and in fact are equivalent. Each has advantages — the infinitary version is conceptually simpler (it has less quantifiers), but the finitary version allows you to ask natural questions — how big does N have to be for this theorem to hold? (This is very much an open problem — even getting good asymptotics on it is open — and is an active area of research.)

Proof of equivalence. The fact that the finitary version implies the infinitary version is relatively straightforward — suppose we assume the finitary statement, and someone gives us a coloring of $\mathbb{N}$ with r colors. Then we only have to look up to the corresponding N, and we can ignore everything else; and we can find a monochromatic solution among them.

The converse is more interesting; the idea is to use a *diagonalization trick*. We assume that Schur's theorem in the infinitary form is true, and we want to show the finitary form is true. Fix r, and assume that the finitary form is false (but the infinitary form is true). Then for every N, there exists some coloring $\varphi_N \colon [N] \to [r]$ avoiding solutions to $x + y = z$. This means we have an infinite sequence of colorings avoiding solutions, and we want to piece them together into a single sequence of $\mathbb{N}$.

We do this using diagonalization, by filtering down to subsequences — we take an infinite subsequence of (φ_N) such that for every k, the color $\varphi_N(k)$ eventually stabilizes. You can do this one at a time (you choose

a subsequence where 1 is always eventually red, then among these you choose one where 2 is eventually blue, and so on). This then gives a coloring $\varphi: \mathbb{N} \to [r]$. This coloring has to avoid monochromatic solutions to $x + y = z$. □

The reason we did this proof is that there'll be many statements (including ones we see today) which have multiple equivalent formulations, infinitary and finitary. It's useful to be able to go back and forth (the infinitary form is a conceptually prettier statement, but the finitary form is what we use to prove the result; it also leads to quantitative questions about how large N has to be).

We'll now prove the statement about FLT mod p.

Proof of $(*)$ *assuming Schur's theorem.* Consider the multiplicative group $(\mathbb{Z}/p\mathbb{Z})^\times$ (which is a cyclic group of order $p-1$), and let $H = \{x^n \mid x \in (\mathbb{Z}/p\mathbb{Z})^\times\}$ be the subgroup consisting of nth powers. Since $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic, H is a cyclic subgroup of order $\gcd(n, p-1) \le n$. The (multiplicative) cosets of H partition $(\mathbb{Z}/p\mathbb{Z})^\times$ into at most n cosets; we can view each of these cosets as a color (so we're coloring $1, \dots, p-1$ using at most n colors).

By (finitary) Schur's theorem, if p is large enoughas a function of n, then there exists a solution to $x+y=z$ viewed as an equation in $\mathbb{Z}$, such that x, y, and z all have the same color.

Since these three numbers have the same color, they lie in the same multiplicative coset; this means there exist X, Y, and Z in $(\mathbb{Z}/p\mathbb{Z})^\times$ such that $x \equiv aX^n$, $y \equiv aY^N$, and $z \equiv aZ^n \pmod p$. This gives $X^n + Y^n \equiv Z^n \pmod p$, giving a nontrivial solution to FLT mod p. □

§1.2 Proof of Schur's theorem

We'll now explain how to prove Schur's theorem. This gets to the heart of the matter of this class and the class in general — the proof will use some graph theory. The graph theory input will be Ramsey's theorem; we'll prove a special case of Ramsey's theorem and use it to prove Schur's theorem.

Theorem 1.4 (Mjulticolored triangle Ramsey)

For every integer r, there exists some N (depending on r) such that if the edges of the complete graph K_N are colored using r colors, then there exists a monochromatic triangle.

The theme of Ramsey theory is that if your system is large enough and you have a finite number of colors, you can always find some large monochromatic orderly structure.

Proof. First let $N_1 = 3$, and define $N_r = r(N_{r-1} - 1) + 2$. We'll show by induction on r that the statement is true for these values of $N = N_r$.

For $r = 1$, this is clearly true. Now suppose we have the claim for $r-1$ colors. Consider an r-edge-coloring of K_N. Pick a vertex v, and consider its neighbors.

Of the $N_r - 1$ outgoing edges from v, by the Pigeonhole principle a lot of them must have the same color — in particular, at least N_{r-1}. Call this color red, and focus on the vertices they go to. We have at least N_{r-1} vertices, so there are two possibilities. Either there exists a red edge between these vertices — in which case we get a red triangle — or there are no red edges — in which case we have one fewer color, and we are done by induction. □

We'll now deduce Schur's theorem from Ramsey's theorem.

Proof of Schur. Given a coloring $\varphi\colon [N] \to [r]$, we want to show that if N is large we can find a monochromatic solution to $x + y = z$. We do this by setting up a graph coloring — color the edges of a complete graph on vertices $\{1, \dots, N+1\}$, where the color assigned to the edge $\{i, j\}$ with $i < j$ is the color of $j - i$ in our given coloring, i.e., $\varphi(j - i)$. This gives a r-coloring of the edges of K_{N+1}.

If N is large enough (with respect to r), then by Ramsey there exists a monochromatic triangle. This means we have three vertices $i < j < k$ such that $\varphi(j - i) = \varphi(k - j) = \varphi(k - i)$. But these three numbers form a solution to $x + y = z$, taking $x = j - i$, $y = k - j$, and $z = k - i$. So this gives a monochromatic solution to $x + y = z$ under ϕ. □

The initial problem was about integers, and we solved this problem by setting up a graph; what did we gain? One way to think about this is that there are techniques we used with graphs that involved picking apart vertices; that would have been very unmotivated if we just worked with the integers. So going from integers to graphs gave us more flexibility, and allowed us to do more arguments that are more natural with graphs than with integers. This will be a common theme in this course — lifting a problem about integers to graphs and using sophisticated tools from graph theory, we can tackle the original problem about the integers.

> **Remark 1.5.** This proof gives some value of $N(r)$; if we work it out, we get $N(r) = \lceil r!\, e \rceil$. (You can roughly see that every time you go up, you multiply by r.) We don't know how tight this bound is; a major conjecture is that $N(r) = C^r$ is enough for some C. (We have no idea how to do this, and it's one of Erdős's favorite problems.) This also relates to several other interesting topics (e.g. Shannon capacity of graphs).

§1.3 History of Additive Combinatorics

We started with Schur's theorem, and it took us to a technique where we used graph theory and Ramsey's theorem to prove Schur's theorem. Historically, this is how the subject started.

Schur's theorem is from 1916; in the following 100 years, this subject grew into additive combinatorics.

§1.3.1 This Class

When Prof. Zhao was a PhD student (at MIT, under Jacob Fox), he learned about a lot of extremal grpha theory; at some point, this led him to learn more things about additive combinatorics and the connection between them. He found this really beautiful (and still works on it). When he started as faculty here, he wanted to distill it into a class. He first taught this class in 2017; part of the goal was to show us what he's learned in his PhD and afterwards, and to show us what are some of the exciting works people are currently doing in modern combinatorics research, so that it's an on-ramp to doing research in this field.

After teaching the course a few times, Prof. Zhao also compiled lecture notes into a textbook that was recently published. He doesn't have a physical copy of the textbook yet, but it is promised to come later this year. A free copy of the book is on Prof. Zhao's website; that will be the primary reference for this class.

On the course homepage, you can find a lot of information about schedule, homework, and class policies.

Homeworks are an integral part of this class; much of the material, you can only really learn by thinking hard about the homework problems. The way they're set up, there are unstarred problems that are somewhat challenging, but also meant to reinforce the material that's learned (you should be able to solve them if you think deeply and understand the material). The starred problems tend to be more challenging, and sometimes require ideas that are not necessarily directly taught in this class (though they should be doable

without additional advanced knwoledge). These problems are somewhat optional, in the following sense (it's the same as 18.226): you are graded on homework, and the determination of your letter grade only comes from the unstarred problems. The grade modifiers +/- are not factored into the official grade; but if you want an A or A+ then you need to solve sufficiently many starred problems.

Prof. Zhao will have office hours, starting next week; in the office hours we can ask Prof. Zhao about the unstarred problems. (We explain what we were thinking, and Prof. Zhao gives us hints.)

There is another component to the assignments, as an experiment this year. At the end of this class, we submit a 2-page open problem proposal; imagine a short document he could reasonably hand to a beginning grad student or advanced undergrad trying to do research. This requires some thought (it's not just copying some problem) — we need to think about the problem a bit and come up with some initial maps, and try to design something that could be fun and realistic for the intended audience. In fact, we could work on this problem (but you don't have to).

This is an important exercise because as students, we are often given research problems to work on; it's an important skill to come up with good problems, and a roadmap or strategy for possibly attacking these problems. More information is on the course homepage.

To supplement that, for the Friday office hours in the math common room, at the same time his research group runs open problem sessions where they present open problems to each other (famous ones, ones they like, ones related to what they're working on); we are welcome to join.

§1.4 Progressions and Additive Combinatorics

Now we'll talk about the history of additive combinatorics, especially regarding to progressions.

Additive combinatorics is a relatively new term, in the sense that you probably wouldn't encounter it before 2000 (previously it was known under different names, such as combinatorial number theory). These types of problems took an exciting turn in around 2000, when there were developments by Gowers, Green, Tao, and so on who inserted a lot of depth to the subject; the term was coined by Tao as a rebranding. It's a very deep and far-reaching subject, and connects many areas of math (including Fourier analysis, graph theory; and others we won't see such as ergodic theory, model theory).

One of the earliest results after Schur's theorem was van der Waerden's theorem, also a foundational result in Ramsey theory:

Theorem 1.6 (van der Waerden's theorem, 1927)

If $\mathbb{N}$ is finitely colored (i.e., colored using finitely many colors), then there exists an arbitrarily long monochromatic arithmetic progression.

We'll use AP to abbreviate arithmetic progression. Schur's theorem was about finding monochromatic solutions to $x + y = z$; here we're finding a different pattern, a sequence a, $a + d$, $a + 2d$, $\ldots$, $a + (k - 1)d$.

The proof of this theorem is involved, but elementary; the earliest involved a color-focusing argument, and the technique is of a similar flavor (though it's much more intricate).

But then people started asking why this result is true — is coloring fundamentally important?

Conjecture 1.7 (Erdős–Turán 1936) — Any subset of $\mathbb{N}$ with positive density contains arbitrarily long arithmetic progressions.

So ErdőS and Turán believed the colors aren't important — if you have 100 colors then one of them occupies 1% of the integers, and this alone should be enough.

Van der Waerden's proof doesn't show anything close to this (it really required colorings).

We won't dwell too much on what positive density means; there are several notions, and it doesn't matter that much which one we take. For concreteness, we use positive *upper density* — if

$$\limsup_{N\to\infty} \frac{|A \cap [N]|}{N} > 0,$$

then we say the set A has positive density.

This conjecture stood for a long time; the first nontrivial case was solved by Roth in the 50's. We can rephrase 'arbitrarily long' as 'contains a k-AP for every k.' Roth's theorem says that the conjecture is true for $k = 3$.

Theorem 1.8 (Roth 1953)
ET is true for $k = 3$.

Roth's proof uses Fourier analysis (also known as the Hardy–Littlewood circle method in analytic number theory).

The full conjecture was resolved a couple of decades later in a landmark paper of Szemerédi.

Theorem 1.9 (Szemerédi 1975)
ET is true.

Roth won a Fields medal (though he had a more famous theorem about diophantine equations); Szemerédi won the Abel prize (he also had a lot more work, but this theorem is really important).

Szermerédi's original proof is relaly complicated; there is a diagram in the introduction explaining the logical flow of all the lemmas and propositions, and this diagram is sufficiently complicated that it takes a minute to figure out that it even leads to the desired conclusion (there are circles everywhere).

A fun story is that Szemerédi didn't even write the paper. Szemerédi has made lots of interesting contributions, but he was also known for not writing his papers; in the acknowledgements he actulaly thanks one of his friends for writing the paper.

This was a combinatorial proof; it's a *tour de force*, and very few people today really understand this proof.

Nevertheless, there was a lot of subsequent effort to find new proofs of this theorem. There's several modern approaches that have opened new areas of math research, all of which are still flourishing today.

- An ergodic theory approach due to Furstenberg — by recasting the problem in terms of ergodic theory, he gave a different proof. At first it wasn't clear whether this proof was basically the same or fundamentally new, but using these ideas people foudn new theorems, and deep extensions of Szemerédi's theorem that to this day we still don't know how to prove using other means; so ergodic theory is now an important part of additive combinatorics, in that it provides insights we would not have otherwise.
- Another approach builds in a sophisticated way on Roth's method. Tim Gowers in the 2000s came up with a pretty revolutionary way to extend Roth's techniques to larger k, starting *higher-order Fourier analysis*. This not only opened new areas of research, but showed us deep insights into the structure of APs; Gowers won the Fields medal partly for this work.
- The third approach, perhaps most relevant to this class, is due to hypergraph regularity; this is credited independently to a team led by Rödl 2005 and independently Gowers. There is another proof of Roth's theorem found by Rusza and Szemeredi that uses graph theory; in a way similar to what we

saw earlier where Ramsey's theorem is useful for proving Schur's theorem, in the 70's they realized there are techniques in graph theory that can be used to prove Roth's theorem. By a lot of ingenous work and deep insights, researchers extended that to hypergraphs, which gives a new combinatorial proof of Szemerédi's theorem. The second chapter of the textbook is on graph regularity, which is an important technique we'll explore in quite a bit of depth.

- THere are also other proofs: a p roof using density Hales–Jewett (which was initially proved using ergodic theory, but a more combinatorial proof was found later in a Polymath project initiated by Gowers).

All these approaches have advantages and disadvantages. The ergodic theory approach allows us to prove more theorems; some of these we now have other proofs (using e.g. hypergraph regularity), but for a lot of them we don't know any other proof. On the other hand, one feature of ergodic theory proofs is that they give no quantitative bounds (there's no function you can write down such that if N is bigger than the function, then the finitary version is true); this is because it involves compactness arguments.

On the other hand, the best quantitative bounds come from Fourier theoretic approaches.

We don't understand the formal links between these approaches, but there are some common themes. One important theme is the *dichotomy between structure and randomness*. We'll see this idea many times throughout this class, and we've probably seen it in other contexts (e.g. signal vs. noise). For example, in graph regularity, we'll see an arbitrary graph can be decomposed into some nice structure plus some pseudorandom components. (Random means we're flipping a coin, and there's a probability distribution. Pseudorandom means something like the prime numbers, which are not necessarily random but exhibit properties or behaviors that are in some specific sense like a random set.)

§1.4.1 Quantitative Bounds

Quantitative bounds are a very active area of research.

Roth's theorem in the infinitary form can be reformulated in the finitary form (it's equivalent in the same sense we saw earlier):

Theorem 1.10 (Roth)
Every 3-AP-free subset of $[N]$ has size at most $o(N)$.

By a quantitative bound, we mean, what functions can we put as the size of the largest 3-AP-free set?

First, what should we expect?

Theorem 1.11 (Behrend 1946)
There exists a 3-AP-free subset of $[N]$ of size at least $Ne^{-c\sqrt{\log N}}$.

We do not know how to improve this construction. (It takes some time to process what this means; you can compare the exponent to $\log n$, and see that $\sqrt{\log N}$ is much smaller. So this is greater than N^{1-c} for every c.)

Roth's proof, using Fourier analysis, gives a bound of

$$O\left(\frac{N}{\log\log N}\right).$$

Since then, there's been a lot of effort to improve this bound; a lot of work and beautiful math has been developed for this.

Until very recently, the best bound was due to BLoom–Sisask 2020, where they broke the logarithmic barrier by proving a bound of $\frac{N}{(\log N)^{1+c}}$ for some $c > 0$. Earlier this year, there was a huge and surprising development by Kelley–Meka 2023, who proved an upper bound basically matching Behrend's upper bound — of $\frac{N}{e^{(\log N)^c}}$ for some small c. You can still argue about the right exponent, but researchers in the field didn't expect anything close to this to come up for a very long time.

For 4-APs, much less is known. THe best result is due to Green–Tao, and has the form $\frac{N}{(\log N)^c}$ for some small c. For longer APs, the best result is due to Gowers, and is of the form $\frac{N}{(\log\log N)^c}$, where the constant c depends on k.

§1.4.2 Extensions of Szemerédi's theorem

Instead of working over $\mathbb{Z}$, we can consider what happens in $\mathbb{Z}^d$. Here the right question is to find *constellations*.

Definition 1.12. Fix a finite set $F \subseteq \mathbb{Z}^d$. We say that A contains an *F-constellation* if there exists some $a \in \mathbb{Z}^d$ and some $t \in \mathbb{Z}_{>0}$ such that A contains $a + tF$.

So we can think of F as some two-dimensional pattern, and we want to find some pattern which is a scaled version of F. The question is, can we always find constellations provided that A has positive density?

The answer is yes; this is the multidimensional Szemerédi theorem.

Theorem 1.13 (multidimensional Szemerédi)
Every subset of $\mathbb{Z}^d$ with positive density contains arbitrary constellations.

Remark 1.14. Prof. Zhao highly encourages us to watch A Beautiful Mind. There is a scene in the movie where Nash and his wife Lisa Nash are at a party, and John Nash tells Lisa, look at the stars, name me a shape (e.g., umbrella), and he finds an umbrella among the stars. Likewise, in a large enough subset of the integer grid,

This was first proved using ergodic theoretic methods; Szemerédi's methods don't seem to be able to prove this. There was a later proof using hypergraph regularity, that also gives a combinatorial proof.

We'll now talk about another extension. First we'll talk about a special case (Furstenberg and Sárközy proved the same result independently, using ergodic theory and Fourier analytic methods respectively).

Theorem 1.15 (Furstenberg–Sárközy)
Any subset of $\mathbb{N}$ with positive density contains two numbers differing by a perfect square.

In other words, we have a pattern $\{x, x + y^2\}$.

Question 1.16. What happens if we allow some other polynomial pattern?

This is also true.

Theorem 1.17 (Polynomial Szemerédi theorem)
If $A \subseteq \mathbb{Z}$ has positive density and $P_1, \ldots, P_k \in \mathbb{Z}[X]$ are polynomials with zero constant term, then there exists $x \in \mathbb{Z}$ and $y \in \mathbb{Z}_{>0}$ such that $x + P_1(y), \ldots, x + P_k(y)$ is contained in A.

This generalizes Szemerédi (by taking the P_i to be linear) as well as Furstenberg–Sárközy. You can also formulate polynomial Szemerédi for the multidimensional statement, and that is also true (if formulated correctly).

The only proof we know right now is due to ergodic theory (though recent work, particularly by Peluse, gives new Fourier analytic proofs for certain settings).

Another important theorem we'll mention is the celebrated Green–Tao theorem:

Theorem 1.18 (Green–Tao theorem)

The primes contain arbitrarily long arithmetic progressions.

They proved this result in 2004, and it was published a few years later; it's one of the celebrated achievements of the century. If there's time, Prof. Zhao will explain the high-level strategies. This is one of Prof. Zhao's favorite theorems; during his PhD he did a lot of extremal graph theory, which ended up leading to some problems related to this theorem and let him understand this theorem better using the perspective from graph theory.

§1.5 What comes next

This was a tour of some of the landmark results in additive combinatorics since Schur's theorem.

Next in the course, we'll look at related problems in graph theory, and explore the subject of *extremal graph theory*. Roth's theorem is about, what is the largest subset of $[N]$ not containing a 3-term arithmetic progression? You can ask similar-sounding problems in the context of graphs — equations may relate to subgraphs, like triangles.

Question 1.19. What is the largest number of edges in a n-vertex triangle-free graph?

We'll answer this question at the start of next lecture. This is an important theorem and one of the foundational results in extremal graph theory, but it's not very hard; we'll see several proofs. However, it's also not a theorem that gives you Roth's theorem, though there seem to be superficial similarities. It turns out there *is* a statement in graph theory that gives Roth's theorem, and it's the one answering the following question:

Question 1.20. What is the maximum numer of edges in a n-vertex graph where every edge is contained in exactly one triangle?

This innocent-sounding question turns out to be quite deep; we don't know the exact answer, but we have some bounds, and even getting any nontrivial result involves sophisticated and interesting methods (graph regularity). It turns out once you know a good bound here, that implies Roth's theorem. It's weird that this statement looks elementary and you can try to play with it using bare hands; but it's actually not so easy.

We'll start by talking about questions of the first type; then we'll develop the tools of graph regularity used to answer the second.

The class is divided into the first half being graph theory and the second additive combinatorics; but there are many related themes and ideas, and Prof. Zhao really wants to convey central connections (using GT to solve problems in AC, as well as common themes like structure vs. randomness).

§2 September 11, 2023 — Forbidding a Subgraph

In this chapter, we're going to look at the problem of forbidding a subgraph. The first instance we'll study today is:

Question 2.1. What is the maximum number of edges in a n-vertex triangle-free graph?

We'll soon see the complete answer and the proof. Then you can imagine, what if instead of forbidding a triangle, we try to forbid some other subgraph? It turns out that there's still a lot of mystery — for some graphs the problem is still open. We'll see what we know and the techniques.

These type of problems are called *Turan problems* (we'll soon see why). First we'll answer this problem.

First, one triangle-free graph is $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ — we take n vertices, split them into two halves (as equally as possible), and draw all edges between them (a complete bipartite graph). This graph has $\lfloor n^2/4 \rfloor$ edges.

This graph is certainly triangle-free, and it turns out that it does have the maximum number of edges; that will be the first fact we prove. (This is the earliest theorem in extremal graph theory, and the start of the subject.)

Theorem 2.2 (Mantel's theorem)
Every n-vertex triangle-free graph has at most $\lfloor n^2/4 \rfloor$ edges.

We'll see a couple of different proofs. All are pretty quick, but they use somewhat different ideas.

Proof 1. Let's start with a n-vertex triangle-free graph $G = (V, E)$, and suppose that it has m edges. Now let's look at an edge xy, and look at the neighborhoods of our two vertices x and y. These two neighborhoods $N(x)$ and $N(y)$ must be disjoint, since G is triangle-free. And if x and y have disjoint neighborhoods, then we must have $\deg(x) + \deg(y) \le n$.

This is true for all edges $xy \in E$, so we can sum this inequality over all edges; this gives us

$$nm \ge \sum_{xy \in E} \deg(x) + \deg(y).$$

What happens if we sum this? Let's look at each term and ask, how many times does some term $\deg(x)$ show up in the summation? It shows up once for every edge that x is involved in, so the answer is $\deg(x)$ many times; this means the sum is equal to

$$\sum_{x \in V} (\deg x)^2.$$

We can then use the Cauchy–Schwarz inequality to lower-bound this quantity (we have a sum of squares, so we can lower-bound it by the square of the sum, appropriately normalized) — we have

$$\sum_{x \in V} (\deg x)^2 \ge \frac{1}{n} \left(\sum_{x \in V} \deg x \right)^2 = \frac{(2m)^2}{n^2}.$$

Rearranging this inequality gives $m \le n^2/4$, and since m is an integer, we get the desired result. □

The key fact we used here (where we used triangle-free) is that given an edge xy, the neighborhoods of x and y are disjoint.

Let's now see a different proof.

Proof 2. In this proof, we're going to start by taking an extremal vertex. Again let G be a n-vertex m-edge triangle-free graph, and let v be a vertex of maximum degree. (If there are multiple, choose one arbitrarily.)

Now look at the neighborhood of v — let A be the neighborhood of v, and call everything else B (so $B = V \setminus A$). Because G is triangle-free, there are no edges in A — so in other words, A is an independent set (otherwise we'd have a triangle). So every edge of the graph has at least one vertex in B.

So if we want to upper-bound the number of edges in the graph, we can sum up all the degrees in B — we have

$$e(G) \le \sum_{x \in B} \deg x.$$

(If we have an edge across the two sets it gets counted once; if we have an edge within B it gets counted twice, so this is certainly an upper bound.)

But we chose v to be a vertex of maximum degree, so each of these terms is at most $\deg v = |A|$. This means

$$e(G) \le \sum_{x \in B} \deg x \le |A|\,|B|\,.$$

Using AM-GM we have

$$|A|\,|B| \le \left(\frac{|A| + |B|}{2}\right)^2 = \frac{n^2}{4}.$$ □

This gives a second proof. Let's look at this proof more closely and ask, when does equality occur (i.e., when do we have exactly $\lfloor n^2/4 \rfloor$ edges)? We won't do this completely on the board, but the point is that we basically want A and B to be as close to each other as possible (or else the AM-GM bound loses too much), and we don't want any edges inside B (or else we lose something in the bound $\deg x \le |A|\,|B|$). So equality holds if and only if the graph is the complete bipartite graph with balanced sets, i.e., $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

Remark 2.3. You can also deduce the equality case from the first proof, though it's a bit harder to see.

Remark 2.4. Why do we not want any edges in B for the equality case? The point is that any edge in B gets counted twice in the summation $\sum_{x \in B} \deg x$, giving extra slack.

That was Mantel's theorem, which gives us the maximum number of edges in a triangle-free graph. So what are the next questions we should ask? Let's replace a triangle with some bigger graph; and one natural graph to ask for is a clique.

Question 2.5. What's the maximum number of edges in a n-vertex K_{r+1}-free graph?

To tell us the answer, we'll first define a Turán graph:

Definition 2.6. The *Turán graph* $T_{n,r}$ is the complete n-vertex r-partite graph with all parts of sizes differing by at most 1.

This is a natural generalization of the balanced complete bipartite graph. For example, $T_{10,3}$ is the graph where we have 3 parts and put the 10 vertices into the three parts as evenly as possible (3, 3, and 4), and put all edges between the parts (but without any edges within a part).

The reason this is called a Turán graph is that Turán proved the following foundational result (because of this theorem and the subsequent work Turán did on the problem, it's called Turán's problem).

Theorem 2.7

The Turán graph $T_{n,r}$ maximizes the number of edges among all n-vertex K_{r+1}-free graphs.

It turns out that it actually *uniquely* maximizes this number of edges.

This is a generalization of Mantel's theorem from avoiding triangles to avoiding cliques of fixed size.

We'll see three or four different proofs. Some of them will be extensions or generalizations of the ideas we saw for Mantel's theorem; some will have different flavors.

Before getting into the proof, we'll analyze the Turán graph a bit more to get some asymptotics to play with.

Fact 2.8 — The number of edges in the Turán graph is

$$e(T_{n,r}) \le \left(1 - \frac{1}{r}\right) \frac{n^2}{2}.$$

There might be an error, but the slack is $O(rn)$ (which is linear in n if we think of r as fixed). So when people say Turán's theorem, they often mean the slightly weaker version with this as the upper bound instead; this is basically as good.

To prove this, first there's an easy observation — r-partite graphs are always K_{r+1}-free. So we can ask, what happens if we restrict ourselves to r-partite graphs? This is a much easier question, because r-partite graphs are very structured. This is pretty easy, but we'll state it:

Lemma 2.9

Among all n-vertex r-partite graphs, $T_{n,r}$ uniquely maximizes the number of edges.

If someone already tells you we're working within the realm of r-partite graphs, then you might as well put in all the edges (so the graph is a complete r-partite); then we need to figure out what part sizes we should choose to maximize the number of edges. The claim states that we should make the parts as close as possible — in other words, we want to show that if there are two parts A and B with $|A| + 2 \le |B|$, then we can get more edges by moving a vertex from B to A. This is pretty easy to show.

Then the name of the game is to reduce the problem of forbidding a clique to getting to a r-partite graph — once we know the structure has to be r-partite, then we are done.

We'll now see several different proofs.

Proof 1. This proof will be an extension of the second proof of Mantel's theorem. That proof began by picking an extremal vertex (one of maximum degree); we'll do the same thing here.

First, we'll do induction on r. The base case $r = 1$ is trivial (if the graph avoids edges, then we have an empty graph); now assume $r > 1$, and that the maximum number of edges in a K_r-free graph is $e(T_{n,r-1})$.

Let $G = (V, E)$ be K_{r+1}-free. As earlier, let v be a maximum-degree vertex. We now have a similar situation as earlier — let A be the neighborhood of v, and let $B = V \setminus A$. (It might be helpful to keep track of a specific value of r, for example $r = 3$ — where we're forbidding a K_4-free graph.)

Because G has no clique on $(r + 1)$ vertices, the set A has no clique on r vertices; this means it is K_r-free.

Similarly to earlier, the number of edges with at least one vertex in B is

$$e(A, B) + e(B) \le \sum_{x \in B} \deg(x) \le |A|\,|B|$$

(again edges in B are counted twice, but this is fine because we want an upper bound; and again $\deg(x) \le |A|$ for all x).

Edges in G can be in one of three places — inside A, inside B, or between A and B. In other words,

$$e(G) = e(A) + e(A, B) + e(B).$$

For edges inside A, we can use the induction hypothesis; and for the remaining terms, we can use the bound derived earlier. This gives

$$e(G) \le e(T_{|A|,r-1}) + |A|\,|B|\,.$$

Finally, we claim that this is at most $e(T_{n,r})$. There are a couple of ways to see this. One straightforward way is that $e(T_{|A|,r-1})$ is some number, and we can calculate it. But another way is — imagine putting $T_{|A|,r-1}$ in A. Then $|A|\,|B|$ is the number of edges we get from putting a complete bipartite graph between A and B. So $e(T_{|A|,r-1}) + |A|\,|B|$ counts the number of edges in some n-vertex r-partite graph, namely the one where we replace A by a Turán graph and put all edges between A and B. And we saw that among all such graphs, the number of edges is bounded by the number of edges in the Turán graph. (This gives a calculation-free way to see the inequality.) □

Remark 2.10. We'll now comment on uniqueness — we want to make sure that every step is tight. For the final step, in order for the inequality $e(A) \le e(T_{|A|,r-1})$ to be tight, we're invoking the induction hypothesis, which is tight only for the Turán graph. For the second bound to be tight, we want to see when the bound $e(A, B) + e(B) \le |A|\,|B|$ is tight; you can check this is only true for a complete bipartite graph. And finally, the final inequality is tight if and only if the numbers of vertices among these r parts are balanced. (These details are fairly straightforward to check.)

That finishes the first proof of Turan's theorem.

Remark 2.11. Before we continue, a few logistical notes.

There will be regular office hours on Friday 4:15–5:15 in the math common room. The format is the following — you are welcome to come and ask about unstarred problems on the problem sets (Prof. Zhao will hear what we've thought and give hints/advice). He will not discuss the starred problems, which are meant to be a challenge. We can also ask about anything related to the topic of the class.

At the same time, he'll run an event with the research group where students present open problems to each other. We are welcome to listen and participate. This is also relevant to one of the assignments discussed, where at the end of the term we turn in a 2-page open problem proposal on a topic related to the topic of this class.

In general Prof. Zhao is happy to stay after class to discuss things (either here or in a different room).

Prof. Zhao is gradually posting problems to the problem set; the first problem set is already partially posted, and when he's done posting all the problems, he will send an announcement to Canvas.

Now let's continue with Turan's theorem. We already saw one proof, and some discussion of the equality cases. In combinatorics, it's often useful to keep track of the equality cases at every step of the proof, because this helps us understand what's going on (if we make a mistake, this can show up). But also, this lecture is fairly exceptional in that we're seeing theorems with very neat, easy-to-describe equality cases. For nearly the entirety of the rest of the course, the theorems we'll encounter have much more asymptotic bounds, where we're so far from knowing equality that even if we get a rough bound, we're very happy. So this lecture is much more exact compared to most of the results we'll see in the rest of this class.

Proof 2. The second proof we'll see starts off somewhat analogously to the first proof of Mantel's theorem we saw at the beginning of the lecture. There we looked at an edge xy, and said that x and y have disjoint neighborhoods. What's the analogy of this for K_4-free graphs? We can instead start with a triangle; then

the three vertices in this triangle should have no common neighbor. That's the starting point of the proof (though after a certain point, it'll diverge from what we saw).

We fix r, and do induction on n. When $n \le r$, the statement is trivial (the clique on n vertices is the answer), so assume that $n > r$, and that the theorem is known for all smaller values of n.

Let $G = (V, E)$ be a n-vertex K_{r+1}-free graph. Furthermore, assume that G already has the maximum possible number of edges among such graphs (so among all n-vertex K_{r+1}-free graphs, we take one with the maximum number of edges).

Claim 2.12 — G already contains K_r as a subgraph.

Proof. Otherwise we could add an arbitrary edge, and G would still satisfy all the same requirements. □

(For example, if we're trying to do K_4-free and our graph doesn't have a triangle, we can throw an arbitrary edge into G and it'll still be K_4-free.)

So the point is that we can already assume G has some copy of a clique on r vertices; let A be a clique on r vertices in G, and let $B = V \setminus A$ be the rest of the graph.

There will be some edges from A to B, but because of what we said earlier — because the graph is K_4-free — these 3 vertices cannot have any *common* neighbors in B. In other words, since G is K_{r+1}-free, every vertex in B has at most $r-1$ neighbors in A.

So then we can compute the number of edges between A and B by summing over all such vertices, and each term is at most $r-1$ — this gives

$$e(A, B) = \sum_{y \in B} \deg(y, A) \le (r-1)(n-r).$$

(Here $\deg(y, A)$ denotes the number of edges from y to A.) Finally, we have

$$e(G) = e(A) + e(A, B) + e(B).$$

We can consider each term separately. Since A is a clique on r vertices, there are $\binom{r}{2}$ edges in A. For the number of edges between A and B, we have the above upper bound. For the number of edges in B, we can use the induction hypothesis. This gives

$$e(G) \le \binom{r}{2} + (r-1)(n-r) + e(T_{n-r}, r).$$

Finally, we claim that this is $e(T_{n,r})$. This can be seen either by direct computation, or more simply by thinking about the equality case — consider the graph $e(T_{n,r})$. If we take out an r-clique from this graph (with one vertex in each part) and use that as our A and think about each individual term, we see that everything adds up correctly. That finishes the proof. □

Remark 2.13. You can also check the equality case using this proof. To check the equality case, we need to see when every inequality is tight. For the inequality $e(A) \le \binom{r}{2}$ to be tight, A has to be a r-clique; that's what we started with. For the inequality $e(A, B) \le (r-1)(n-r)$ to be true, every vertex in B must be adjacent to $r-1$ vertices in A. For $e(B) \le e(T_{n-r,r})$ to be tight, B has to be the Turán graph; in particular it has to be r-partite. You can show that this forces the entire graph to be r-partite.

Now let's look at oen more proof.

Proof 3. Here we'll use a method called *Zykov symmetrization*. The idea is that if you start with a graph that's not the Turán graph, we should be able to make some local changes (namely, symmetrizations) to make the graph even better.

As before, let G be a n-vertex K_{r+1}-free graph with the maximum possible number of edges.

> **Claim 2.14 —** If x and y are non-adjacent vertices, then $\deg(x) = \deg(y)$.

Proof. Suppose otherwise — so x and y are non-adjacent, but $\deg(x) > \deg(y)$. So we have a garph with x and y, and some vertices coming out of them (some of which are shared); and a bunch of stuff possibly happening between those vertices.

Then what we can do is transform this graph by deleting y and replacing it by a clone x' of x (a new vertex whose neighborhood is exactly the same as x). We claim that:

- If we start with a graph that's K_{r+1}-free, then the graph remains K_{r+1}-free. This is because a new clique would have to involve the new vertex x'; and it must also involve x, since otherwise we'd see it from x in the original graph. But there's no edge between x and x'.
- This graph has strictly more edges — this is because we take a vertex with $\deg(y)$ edges, and replace it with one with $\deg(x) > \deg(y)$ edges.

So since G has the maximum possible number of edges, it can't be possible to do this; this means all non-adjacent vertices have the same degree. □

> **Claim 2.15 —** Suppose that x is not adjacent to y, adn not adjacent to z. Then y is not adjacent to z.

In other words, non-adjacence is transitive.

Proof. Consider three vertices such that x is not adjacent to y and not adjacent to z. Then these three vertices have the same degrees d by the earlier claim.

Suppose that y and z were adjacent to each other; then we have d edges coming out of x, $d-1$ coming out of y (other than the edge yz), and $d-1$ out of z.

Now consider the operation where we delete y and z, and clone x twice (creating x' and x'').

We can see that the graph remains K_{r+1}-free, and that the number of edges has to increase (by exactly 1 — we lost the edge yz and gained two more). That finishes the claim. □

So we've learned that in our extremal graph G, non-adjacency is an equivalence relation (namely, it satisfies this transitivity property). This means the complement of the graph must consist of disjoint cliques; in other words, G is a complete multipartite graph. The number of parts cannot be more than r (or else we'd have a clique of size $r+1$), so the graph is r-partite (it's possible one part has size 0, in which case the number of edges is not optimal); and then by the earlier lemma we're done. □

> **Remark 2.16.** It's also easy to check the equality case for this proof.

> **Remark 2.17.** The structure of the proof is that we start with a G that's extremal, and then we prove two claims that tell us something about the property of G — in particular, that non-adjacency in G is an equivalence relation. So if we look at the complement of G, all the connected components must be cliques, because adjacency is an equivalence relation; so the complement of G consists of disjoint cliques, which means G itself is complete multipartite.

We have time to see one more proof. This proof has a completely different flavor — it's a probabilistic proof. It'll show a slightly weaker claim, but this is essentially good enough. We'll prove Turán in the following (slightly weaker) form:

Theorem 2.18

The number of edges in a n-vertex triangle-free graph is at most

$$\left(1-\frac{1}{r}\right)\frac{n^2}{2}.$$

This is the same asymptotic as the number of edges in $T_{n,r}$, and for most purposes it's just as good.

Proof. Let G be a n-vertex K_{r+1}-free graph. Consider a uniform random ordering of vertices, and let

$$X = \{v \mid v \text{ is adjacent to all earlier vertices in the random order}\}.$$

So X is some subset of vertices; it's a random subset (since it depends on our random order), and X is always a clique (every vertex of X is adjacent to all its predecessors).

Let's compute the probability that a specific (fixed) vertex $v \in V$ is in X. This is the probability that v appears before all of its non-neighbors in the random ordering; this means it appears first among all its non-neighbors (along with itself), so

$$\mathbb{P}(v \in X) = \frac{1}{n-\deg(v)}.$$

(There are $n-\deg(v)$ candidates it's competing with, including itself — if we permute $n-\deg(v)$ things, the probability a specific one comes first is $\frac{1}{n-\deg(v)}$.)

But on the other hand, we have $|X| \le r$ for all X (since the graph is free of cliques of size bigger than r). So in particular $\mathbb{E}[|X|] \le r$. But on the other hand, by linearity of expectation we have

$$\mathbb{E}[|X|] = \sum_{v\in V}\mathbb{P}(v\in X) = \sum_{v\in V}\frac{1}{n-\deg(v)}.$$

By convexity (putting the average into the denominator), we get

$$\mathbb{E}[|X|] \ge \frac{n}{n-\frac{1}{n}\sum_{v\in V}\deg v} = \frac{n}{n-\frac{2m}{n}},$$

and rearranging gives $m \le (1-1/r)n^2/2$. □

§2.1 A Preview

Today we saw two proofs of Mantel's theorem and four of Turán's theorem. In the homework, there will be a few problems asking us to modify these proofs to the problems being asked. We'll now see a preview of what's to come in the rest of the chapter.

We can reformulate these problems in terms of studying the extremal number of H-free graphs:

Definition 2.19. We define $\mathrm{ex}(n, H)$ as the maximum number of edges in a H-free n-vertex graph.

We saw what happens when H is a triangle or a clique, but we can ask what happens for any graph H.

A subtle but important point is what it means for a graph to be H-free — we define this to mean that the graph has no *subgraph* isomorphic to H.

It's important to distinguish the notion of *subgraph* vs. *induced subgraph*. A graph is a set of vertices and a set of edges. A *subgraph* means we're allowed to delete vertices and edges, while an *induced subgraph* means we're only allowed to delete vertices (but not edges). In other words, an induced subgraph means we select some vertices and keep all edges between them.

With very few exceptions, when we talk about H-free and subgraphs we mean subgraphs (when we want induced subgraphs, we will explicitly say this).

What we saw is the following: Mantel's theorem tells us that

$$\operatorname{ex}(n, K_3) = \left\lfloor \frac{n^2}{2} \right\rfloor,$$

and Turán's theorem tells us that

$$\operatorname{ex}(n, K_{r+1}) = e(T_{n,r}) = \left(1 - \frac{1}{r} + o(1)\right) \frac{n^2}{2}$$

(in the asymptotic r is fixed and $n \to \infty$).

There are a couple of directions to go from here.

Question 2.20. Can we get a result like this for a general graph H?

The answer is yes:

Theorem 2.21 (Erdős–Stone–Simonovits)
For fixed H as $n \to \infty$, we have

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

Here $\chi(H)$ is the chromatic number.

This completely answers the question in a way that generalizes Mantel's theorem and Turán's theorem; we'll see a proof in the next lecture.

However, it's also unsatisfactory in the following sense: when $\chi(H) = 2$, i.e., H is bipartite, this theorem simply tells us that $\operatorname{ex}(n, H) = o(n^2)$. That's something, but we would like to get a more precise answer. And for a very small number of graphs, we actually do know what the right answer is (asymptotically, meaning we know the right rate of growth) — it'll be some exponent between 1 and 2. But for most bipartite graphs H, this is still open.

Next lecture we'll prove the Erdős–Stone–Simonovits theorem. But before we finish the class, this is a subject with very deep Hungarian roots (because of the influence of Paul Erdős — if we haven't heard of him we should look him up). He's a very important figure in math; he's written over 1600 papers, and is very prolific and well-collaborated. A lot of combinatorics, extremal combinatorics, and extremal graph theory has ties to him; and he's influenced many Hungarian mathematicians. So we'll see a quick crash course on pronouncing Hungarian names.

If you see the letter 's' (e.g. Erdős), the closest English pronounciation is /sh/. Meanwhile, if you see 'sz' (e.g. Szemerédi) this is close to /s/. Stone is an English mathematician, so this doesn't apply.

§3 September 13, 2023

Last time, we proved Mantel's theorem and Turán's theorem. More generally, we introduced the Turán problem:

Definition 3.1. For a graph H, we define $\mathrm{ex}(n, H)$ as the maximum number of edges in a n-vertex H-free graph.

Last time we determined this number exactly when H is a clique. Today we'll look at other graphs H and better understand this quantity.

§3.1 The bipartite case

We'll start with the case where $H = K_{s,t}$ is a complete bipartite graph. This problem is generally known as the Zarankiewciz problem — to determine $\mathrm{ex}(n, K_{s,t})$. (The original problem was about $\{0,1\}$-integer matrices, but it's asymptotically the same.)

We have rather poor understanding of this problem — it's nothing like last time, where there's a precise answer. We don't even know the right asymptotic, except in a small number of cases. Still, there's a lot of interesting mathematics.

The main result is the following:

Theorem 3.2 (Kővári–Sós–Turán)
For all $s \le t$, there exists a constant $C = C(s,t)$ such that

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

For example, this gives you an upper bound on the Turán number for $K_{2,2}$.

Remark 3.3. This theorem can also be abbreviated to the KST theorem, which conveniently is also the object of study.

Proof. We'll count in two ways the number of stars — let G be a n-vertex m-edge $K_{s,t}$-free graph. Let X be the number of copies of $K_{s,1}$ in G; we'll count X in two ways, which give an upper and lower bound, and by comparing them we'll get this claim.

The two ways roughly correspond to starting from the left (embedding s vertices and a common neighbor), and starting from the right (embedding a single vertex and choosing s neighbors).

We first get an upper bound on X. Since G is $K_{s,t}$-free, every set of s vertices have at most $(t-1)$ common neighbors — or else t of the common neighbors together with s would form a $K_{s,t}$. This gives that

$$X \le \binom{n}{s}(t-1)$$

(we can choose the s vertices on the left and embed them in at most $\binom{n}{s}$ ways; they have at most $t-1$ common neighbors, so there are at most $t-1$ ways to extend them into a star).

Now let's get a lower bound. For each vertex $v \in V(G)$, there are exactly $\binom{\deg v}{s}$ ways to extend v into a star (we need to choose s of its neighbors). So

$$X = \sum_{v \in V} \binom{\deg v}{s}.$$

We want to lower-bound this quantity; roughly speaking, if we view s as a constant, then this binomial coefficient is a convex function in the degree. To be more precise, the function $\binom{x}{s}$ is not quite convex, but

we can modify it to be convex — define

$$f_s(x) = \begin{cases} \frac{x(x-1)\cdots(x-s+1)}{s!} & \text{if } x \geq s-1 \\ 0 & \text{otherwise.} \end{cases}$$

(The binomial coefficient is convex past $s-1$ but goes up and down through the zeros, so we just flatten it out.) This function is indeed convex, and $f_s(x) = \binom{x}{s}$ for all nonnegative integers x. So we have

$$X = \sum_v f_s(\deg v),$$

and using convexity we can lower-bound this by

$$X \geq n f_s\left(\frac{2m}{n}\right)$$

(where $2m/n$ is the avertage degree). Putting these two things together, we get

$$n f_s\left(\frac{2m}{n}\right) \leq X \leq \binom{n}{s}(t-1).$$

Here s is constant, so roughly speaking, the left-hand side is

$$(1+o(1))\frac{n}{s!}\left(\frac{2m}{n}\right)^s,$$

and the right-hand side is

$$(1+o(1))\frac{n^s}{s!}(t-1).$$

Rearranging gives that

$$m \leq \left(\frac{(t-1)^{1/s}}{2} + o(1)\right) n^{2-1/s}. \qquad \square$$

This gives us an upper bound on the number of $K_{s,t}$. There are several lingering questions — how tight is this bound? Is this asymptotically close to optimal, or is there some way to improve this exponent? That's a major open problem.

Conjecture 3.4 — The KST bound is tight — namely, all $s \leq t$, there exists a constant $c = c(s,t)$ such that for all $n \geq 2$, we have

$$\operatorname{ex}(n, K_{s,t}) \geq c n^{2-1/s}.$$

In other words, the conjecture is that there's a matching lower bound to KST. (Prof. Zhao likes to use C for big constants and c for small constants; it's a nice convention.)

This conjecture is very much open, except for a few cases where we do know it's true (which we'll see in the next lectures) — we know it's true for $s = 2$ and $s = 3$, and when t is sufficiently large with respect to s (i.e., $t > t_0(s)$ for some function t_0). For all other cases, the conjecture is open — in particular, it is still open for $K_{4,4}$. There are even doubts about whether the conjecture is true for $K_{4,4}$. The fundamental reason it's open is we don't have good ways to construct graphs that are $K_{4,4}$-free and have lots of edges — in the next couple of lectures we'll see some techniques for lower-bound constructions (all we've talked about here is upper-bounds, but the other side of the question is how to construct graphs that are H-free but have lots of edges).

A corollary of KST is the following:

Corollary 3.5

For every bipartite H, there exists a constant $c = c_H > 0$ such that

$$\mathrm{ex}(n, H) = O_H(n^{2-c}).$$

Proof. Given H, we can embed it in some $K_{s,t}$ (it's a bipartite graph, so we can fill in some missing edges to get a complete one). If a graph is H-free then it is automatically $K_{s,t}$-free (if it contains $K_{s,t}$ then it contains a copy of H), so

$$\mathrm{ex}(n, H) \le \mathrm{ex}(n, K_{s,t}) \lesssim_s n^{2-1/s}.$$ □

For most problems, this upper bound is not tight and can be improved; but there may be some cases where it is tight.

This is the story so far for bipartite graphs; there's still a lot we don't know. We'll now move on to non-bipartite graphs; here we have a much more satisfactory understanding.

§3.2 The non-bipartite case

For non-bipartite graphs, at the end of last lecture we mentioned the following theorem:

Theorem 3.6 (Erdős–Stone–Simonovits)

For all H, we have

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

This is interesting because it says $\mathrm{ex}(H)$ only really depends on the chromatic number.

Definition 3.7. The *chromatic number* $\chi(H)$ is the minimum number of colors needed to properly color H — i.e., to color the vertices of H such that no two adjacent vertices receive the same color.

The chromatic number of a triangle is 3. But you could have much more complicated graphs (e.g. the Petersen graph — which has 10 vertices) that also have chromatic number 3 (the Petersen graph is a nice graph that's a counterexample to a lot of statements) — it isn't 2 because there's a 5-cycle, but there is a way to 3-color the vertices.

So we have a fairly complicated graph; nevertheless, via the Erdős–Stone–Simonovits theorem, we know that the extremal number of the Petersen graph is the same (asymptotically) as the extremal number of a triangle, namely

$$\left(\frac{1}{4} + o(1)\right) n^2.$$

This is kind of amazing, since it seems to be a much more complicated object.

The rest of this lecture will be focused on proving this theorem; we'll see the proof up to some small details (which we can think about on our own).

Remark 3.8. This theorem applies to all graphs, not just bipartite ones. But when H is bipartite — equivalently, $\chi(H) = 2$ — this tells us that $\mathrm{ex}(Hn, H) = o(n^2)$. This is some information, but it's not the end of the story — we want to get better asymptotics.

To prepare ourselves for this proof, let's try to better understand the structure of the extremal numbers. Let's think about this problem at a more abstract level first, before diving into the specifics.

If I give you a graph G, I can talk about its edge density. There are several ways to define edge density, similar to each other but with small nuances; the notion we'll use for this lecture is that the edge density is

$$\frac{e(G)}{\binom{v(G)}{2}}$$

(the fraction of all possible edges we have). For this definition of edge density, we have monotonicity of extremal numbers:

Lemma 3.9

We have

$$\frac{\mathrm{ex}(n+1, H)}{\binom{n+1}{2}} \le \frac{\mathrm{ex(n,\ H)}}{\binom{n}{2}}.$$

In other words, the extremal number viewed as a density is monotone decreasing as n increases.

Proof. Suppose we have a graph G on $n+1$ vertices that is H-free. Let $S \subseteq V(G)$ be a subset obtained by deleting one of the vertices, chosen uniformly at random (and keeping the other n).

Now consider the induced subgraph $G[S]$. This graph has n vertices, and is also H-free; so its number of edges is $e(G[s]) \le \mathrm{ex}(n, H)$. Dividing both sides by $\binom{n}{2}$, we have

$$\frac{e(G[S])}{\binom{n}{2}} \le \frac{\mathrm{ex}(n, H)}{\binom{n}{2}}.$$

Now for random S, what's the expectation of the left-hand side? The left-hand side is the edge density of the induced subgraph, and as we vary over all induced subgraphs, by linearity of expectation we'll just get the overall edge density — so

$$\mathbb{E}[\mathrm{LHS}] = \frac{e(G)}{\binom{n+1}{2}}.$$

This gives the desired claim. □

We've just seen that $\mathrm{ex}(n, H)/\binom{n}{2}$ is monotone decreasing, so it has a limit.

Definition 3.10. The *Turán density* $\pi(H)$ is defined as

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n, H)}{\binom{n}{2}}.$$

This is some well-defined number — the limit edge-density of a H-free graph (we only care about n large for this lecture). There are a few different, equivalent, ways to interpret this quantity. In particular, $\pi(H)$ is the smallest real number such that the following is true: for every $\varepsilon > 0$, as long as n is large enough (with respect to ε), every n-vertex graph with at least $(\pi(H) + \varepsilon)\binom{n}{2}$ edges contains a copy of H.

Example 3.11

Mantel's theorem gives that $\mathrm{ex}(n, K_3) = \lfloor n^2/4 \rfloor$, which means that

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

(In the extremal example — the complete bipartite graph — half of all possible edges are present.) Likewise, Turán's theorem implies that

$$\pi(K_{r+1}) = 1 - \frac{1}{r}.$$

The ESS theorem says that

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

That in some sense puts an end to the story, because it determines this quantity for all graphs. So in this sense, once we prove this theorem it finishes the story for all non-bipartite graphs.

But this is in some sense a happy coincidence — in *graph* theory we know the answer, but even if you change the problem a little bit, we don't know the answer. We can ask the same question for hypergraphs instead — where the edges are *triples* of vertices, rather than pairs. We can also define the Hypergraph Turán number; then we can ask, what's the hypergraph Turán number of the tetrahedron $K_4^{(3)}$ (the exponent denotes 3-uniform)? We don't know the answer — we don't know $\pi(K_4^{(3)})$, and though there are conjectures, we're far from proving them.

§3.3 Proof of Erdős–Stone–Simonovits

In the rest of the lecture, we'll prove the Erdős–Stone–Simonovits theorem, that

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

There's an easy way to prove a lower bound, using the same construction as the Turán graph construction:

Proof of lower bound. If $\chi(H) = r + 1$, then $T_{n,r}$ is H-free — it's an r-partite graph with r 'islands', and we can't embed H into it because H isn't r-colorable. So then

$$\mathrm{ex}(n, H) \geq e(T_{n,r}) = \left(1 - \frac{1}{r} + o(1)\right) \frac{n^2}{2}. \qquad \square$$

(Asymptotically this can't be improved, but you can actually do better with lower-order terms; we'll see that on a problem set.)

§3.3.1 Supersaturation

We need some additional ideas for the proof of the upper bound. One has the nice name of *supersaturation*. On the homework, some of the problems have the following flavor: Mantel's theorem tells us a triangle-free graph has at most $n^2/4$ edges. What if you have slightly more than $n^2/4$ edges? Then you'll have some triangles; but the point of supersaturation is that you actually have a *lot* of triangles. (This will be some rough statement that's quite general; but for specific situations where you want tighter bounds, e.g. on the pset, you need additional ideas.)

Lemma 3.12

For all $\varepsilon > 0$ and H, there exists some $\delta > 0$ and n_0 such that for every graph G on $n \geq n_0$ vertices with at least $(\pi(H) + \varepsilon)\binom{n}{2}$ edges, G must contain at least $\delta n^{v(H)}$ copies of H¿

The point here is that if G is above the Turán threshold by some significant amount, then G contains lots of copies of H.

What's going on here? Imagine H is a triangle, so $\pi(H) = \frac{1}{2}$; Mantel's theorem tells us that the maximum density is $\frac{1}{2}$ in a triangle-free graph. This tells us that if G has edge density even 0.51, then it has a number of triangles on the order of n^3 (which is the right order of magnitude for counting triangles, since there are 3 vertices).

Proof. The proof is via a sampling argument (which is elementary but important). By the definition of Turán density, there exists some n_0 (which may depend on H and ε) such that every graph with n_0 vertices and at least $(\pi(H)+\frac{\varepsilon}{2})\binom{n_0}{2}$ edges contains H as a subgraph. (This is simply a reformulation of the definition of Turán density — if your edge density is over this threshold, then as long as n is large enough, we have H as a subgraph.) Fix some such n_0.

Now let $n \geq n_0$, and suppose that G is a n-vertex graph with at least $(\pi(H)+\varepsilon)\binom{n}{2}$ edges, as in the statement we're trying to prove. Let $S \in \binom{V(G)}{n_0}$ be a random subset of $V(G)$ of size exactly n_0, chosen uniformly at random, and consider the induced subgraph $G[S]$ (so we pick n_0 random vertices and look at the subgraph they induce). Let

$$X = \frac{e(G[S])}{\binom{n}{2}}$$

denote the edge density of this subgraph. (We saw a similar argument earlier when proving monotonicity.) By averaging (as earlier), the expected value of X is equal to the edge density of G (every edge gets hit an equal amount), so

$$\mathbb{E}[X] = \frac{e(G)}{\binom{n}{2}} \geq \pi(H) + \varepsilon.$$

Now X is some random variable between 0 and 1, so if it is bigger than some quantity, then we know that it must be quite a bit larger than $\pi(H)$ for some not-small amount of probability — more specifically, this statement implies that $X \geq \pi(H) + \frac{\varepsilon}{2}$ with probability at least $\frac{\varepsilon}{2}$, i.e.,

$$\mathbb{P}\left(X \geq \pi(H) + \frac{\varepsilon}{2}\right) \geq \frac{\varepsilon}{2}.$$

(This is a common step in such arguments; if this were false, we could bound $\mathbb{E}[X]$ to get some number strictly less than $\pi(H) + \varepsilon$.)

Then with probability at least $\frac{\varepsilon}{2}$, we have that $G[S]$ has edge density at least $\pi(H) + \frac{\varepsilon}{2}$, and therefore has a copy of H; and therefore G also has a copy of H.

So as S varies over all $\binom{n}{n_0}$ subsets, this gives us at least $\frac{\varepsilon}{2}\binom{n}{n_0}$ copies of H. But we've overcounted — some copies of H might get counted many times by overlapping subsets. So we need to estimate the number of overcounts — each copy of H is counted at most $\binom{n-v(H)}{n_0-v(H)}$ times, and putting these together gives that the number of copies of H is at least

$$\frac{\frac{\varepsilon}{2}\binom{n}{n_0}}{\binom{n-v(H)}{n_0-v(H)}} \geq \delta n^{v(H)},$$

where δ depends on all these constants. □

§3.4 Proof of ESS

How does supersaturation help? We'll see that in a second.

The naming is interesting because there were a couple of different papers — one by Erdős–Stone and one by Erdős–Simonovits. The most important result is from the Erdős–Stone paper, and is as follows:

Theorem 3.13 (Erdős–Stone)
If $H = K_{s,\dots,s}$ is the complete r-partite graph with s vertices in each part, then

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

The point is if we start with some H with chromatic number r, we have r islands with some edges going between them; and we might as well put in all of them. So the original statement follows from this one (it suffices to consider the case where H is a complete multipartite graph).

We'll now introduce the concept of a blowup — if we start with a graph H, we clone each vertex some number of times, and replace every edge by a complete bipartite graph between the clones of its endpoints.

Definition 3.14. The r-blowup $H[r]$ is the graph where we repeat each vertex of H r times, and draw a complete bipartite graph between the clones of the endpoints of each edge.

Then Erdős–Stone says that $\pi(K_r[s]) = 1 - \frac{1}{r-1}$ — the complete r-partite graph $H = K_{s,\dots,s}$ is the s-blowup of K_r.

We'll prove this by rephrasing the problem in terms of r-uniform hypergraphs.

Definition 3.15. A *r-uniform hypergraph* (or *r-graph*) consists of vertices V and edges E, where each element of E is an r-element subset of V.

So usual graphs are 2-graphs; we'll be looking at r-graphs.

Likewise, we can define $\operatorname{ex}(n, H)$ for a r-graph H to be the maximum number of edges in an n-vertex H-free r-graph (the notion of subgraphs in hypergraphs is a straightforward generalization of that for graphs). As mentioned earlier, even for the generalization of triangles to simplices we don't know the answer; but we can still define it.

At the begining of this lecture, we saw the KST theorem, which says that

$$\operatorname{ex}(n, K_{s,t}) = O_{s,t}(n^{2-1/s}).$$

We're going to prove Erdős–Stone by developing a hypergraph extension of KST.

Lemma 3.16 (Hypergraph KST)
For all fixed $r \geq 2$ and s, we have

$$\operatorname{ex}(n, K^{(r)}_{\underbrace{s,\dots,s}_{r}}) = o(n^r).$$

Here this graph is the r-uniform hypergraph where we first draw r vertex sets, each with s vertices, and then put down as edges all the r-vertex subsets that cut across all the parts. (This is the natural generalization of a complete bipartite graph.) It is called the *complete r-partite r-graph*.

Remark 3.17. If we follow the proof, we can get that the exponent is r minus some small constant; but for the purposes of this proof, $o(n^r)$ is good enough.

Example 3.18

For example, for 3-graphs this tells us that $\mathrm{ex}(n, K^{(3)}_{s,s,s}) = o(n^3)$ — if we try to avoid this graph in a 3-graph, then we have edge density asymptotically 0.

We won't dwell too much on the proof; after a couple of things, we can basically extend the earlier proof of KST, with a few modifications.

Proof sketch for 3*-graph KST.* In the proof of KST for graphs, we counted stars $K_{s,1}$ in two ways. Here we'll count something similar — let X denote the number of copies of $K^{(3)}_{s,1,1}$, the natural generalization of a star (tri-partite 3-graphs with s vertices in one part, and one vertex in the other two). We'll count X in two ways.

For the upper bound, we start by embedding the s vertices — given $S \in \binom{V}{s}$, we can look at how many ways there are to extend it into this graph. Let T be the set of unordered pairs that form $K^{(3)}_{s,1,1}$ with S (this will be an analog of common neighbors, but instead of vertices we're looking at unordered pairs).

Then we can view T as a graph — T consists of unordered pairs. Then T as a graph is $K_{s,s}$-free — if T has some $K_{s,s}$, then we can combine it with S to get a $K_{s,s,s}$.

So now we can apply the KST theorem to T; this gets us that

$$X \lesssim \binom{n}{s} n^{2-1/s}$$

(since there are at most $n^{2-1/s}$ edges in this graph, so at most this many ways of picking an unordered pair in T).

For the lower bound on X, we start by choosing the pair (of vertices in the 1-parts) — write $\deg(u, v)$ to be the number of triples (edges) in G containing both u and v. Then

$$X = \sum_{u,v} \binom{\deg(u,v)}{s}$$

(similarly to how when counting stars, we fixed a vertex and counted choices of s of its neighbors). As earlier, we can lower-bound this using a convexity argument; and combining the bounds and comparing gives the conclusion. □

Here we proved 3-uniform KST using 2-uniform KST; more generally, you can induct on r (you prove 4-uniform KST using 3-uniform KST, and so on).

With that out of the way, we'll prove the Erdős–Stone theorem. We'll first recast this problem in terms of the Turán problem for an r-graph.

Proof. Let G be a H-free graph on n vertices, where H is the complete r-partite graph with s vertices in each part.

We then construct an auxiliary r-graph $G^{(r)}$ in the following way: this r-graph will have the same vertex set as G, and its edges will be precisely the r-cliques in G. (If $r = 3$, then every time we see a triangle in G, we put in a triple into the 3-uniform hypergraph — when talking about 3-graphs we'll say *triple* to refer to the edges.)

Note that since G is $K_{s,\dots,s}$-free, then $G^{(r)}$ must be $K^{(r)}_{s,\dots,s}$-free (with r copies of s). (It might be helpful to keep in mind $s = 2$ and $r = 3$ as an example — we have a graph, and a 3-graph on top of it consisting of all the triangles.) Then by hypergraph KST, we know that

$$e(G^{(r)}) = o(n^r).$$

What does that mean for the original graph? Translating back to the original graph, G has $o(n^r)$ copies of K_r.

And now we're back to supersaturation — G has very few copies of K_r. That means the number of edges cannot be significantly above the Turán threshold — by supersaturation, we must have

$$e(G) \le (\pi(K_r) + o(1))\binom{n}{2}.$$

(If this were not true — $e(G)$ were significantly above $\pi(K_r)$ — then supersaturation tells us G must have lots of copies of K_r, and this is not the case.)

But we also know by Turán's theorem that $\pi(K_r) = 1 - \frac{1}{r-1}$; that finishes the proof. □

Let's synthesize what happened. The statement of ESS tells us that the extremal number is entirely determined by the quadratic number (at least, for the first-order term). The lower bound comes from the Turán graph construction. For the upper bound, we first reduce to the case of a complete r-partite graph. At the moral level, what we're saying is that if we can understand some object, then we can also understand something related to its blowup (going from K_r to $K_r[s]$). The way we do this is by starting with H, and building an auxiliary r-graph whose edges are the r-cliques of G. Originally G was free of something, and that translates to G_r being free of something. Now we're in the r-partite case of r-graphs, which we can handle by the extension of KST; that gives us a bound on the number of edges in the r-graph. Then we use supersaturation to go back to G and deduce a statement about the edge density in G.

> **Remark 3.19.** The goal was to prove the extremal number for the blow-up; the way this came up is that if we start with arbitrary H of chromatic number 3, that's contained in the blowup of a triangle. We didn't really use the blowup notation, but these are the objects we're talking about.

§4 September 18, 2023

§4.1 A geometric application of KST

Last time, we discussed the KST theorem:

> **Theorem 4.1** (KST)
> $\mathrm{ex}(n, K_{s,t}) = O_{s,t}(n^{2-1/s})$.

Here we're looking at the maximum number of edges in a n-vertex $K_{s,t}$-free graph.

The following is a classic open problem:

> **Question 4.2** (Erdős unit distance problem)**.** What is the maximum number of unit distances formed by n points in the plane?

You get to put down n points in the plane, and we want to maximze the number of pairs of them that form unit distances.

Example 4.3

If $n = 3$, we can put our points on the vertices of a unit equilateral triangle, giving 3.

If $n = 4$, we can take a 60° rhombus, giving 5 pairs; it's not possible to get all pairs.

If $n = 5$, the best we can do consists of three equilateral triangles in alternating orientation, to get a trapezoid.

If $n = 6$, it's optimal to take two unit triangles with unit translation between them.

If $n = 7$, the best we can do is a hexagon with its center.

When n gets large, there's no discernable pattern.

What's an obvious construction for n? One is to chain up n points along a line; that's certainly one way to get n points, and $n - 1$ unit distances.

Are there better ways? Erdős considered the following construction — what if we put these points on a square grid? Imagine for simplicity that n is a perfect square (if n is not a perfect square, you can do this approximately) — and put the n points in a square grid. We want our unit distance to be *not* the grid length, but some distance which occurs many times in the square grid. In other words, we choose some r so that $r^2 = a^2 + b^2$ for many values of pairs $(a, b) \in [\sqrt{k}]^2$, so that this length comes up many times.

We can do some number theory to choose a value of r; it turns out that taking products of primes is good, and by using various estimates from number theory, Erdős obtained that by choosing r optimally (where r is what we consider to be the unit distance), we can get at least

$$n^{1+c/\log\log n}$$

unit distances.

This is some number, that is superlinear — you can check that this grows faster than linear. But it's not that much faster than linear.

Conjecture 4.4 (Erdős unit distance conjecture) — n points in $\mathbb{R}^2$ form at most $n^{1+o(1)}$ unit distances.

So it's not possible to arrange n points in $\mathbb{R}^2$ to get $n^{1.01}$ unit distances, for large n. This is a major open problem, and we're very far from getting anything close to this answer. But today we'll see an upper bound of the form $n^{3/2}$.

Remark 4.5. This is related to the *distinct distances* problem, which was solved by Guth and Katz; we'll comment on that later.

The point is that you can think of n points as n vertices, and form a unit distance graph (where the points are vertices and edges are unit distances).

Claim 4.6 — The *unit distance graph* is $K_{2,3}$-free.

Proof. Assume that we have two points, and they ahve three common neighbors. What does it mean to be a common neighbor? It has to be unit distance to both points. This means it has to lie on the unit circle centered at the left point, as well as the unit circle centered at the right point. But these two circles intersect in at most two points, contradiction. □

Then by KST, we have $\mathrm{ex}(n, K_{2,3}) = O(n^{3/2})$. This means the unit distance graph has atmost $n^{2/3}$ edges, proving the upper bound.

This is not the best that we know. Later on in the course, we'll see another very short proof using more information about the topology of the plane (in particular, a theorem by Szemerédi about point-line incidences); this will give a bound of $O(n^{4/3})$. This remains the best bound known. There are reasons that it's very hard to improve this bound — if you change the problem from circles to parabolas, then this bound is tight. So there has to be some combinatorics that distinguishes circles from parabolas, which is part of the reason this is difficult.

Remark 4.7. Has the unit distance problem been solved in higher dimensions? It's also interesting in 3 dimensions. But in 4 dimensions, the problem is very easy — the trivial bound of $O(n^2)$ holds, because we can have two orthogonal 2-planes, and then we can put two circles on these two planes which are orthogonal to each other; then every point on the first circle has unit distance to every point on the second.

There was a question about the relationship between this problem and another important Erdős problem:

Question 4.8 (Erdős distinct distances problem)**.** What's the minimum number of distinct distances formed by n points in the plane?

If we take n generic points, they form quadratically many distances. If you put the n points on a line, you get around n distinct distances. You can do better than this — the square grid construction gives

$$\Theta\left(\frac{n}{\sqrt{\log n}}\right)$$

(this requires some calculations from analytic number theory). Erdős conjectured that this is optimal.

Like the unit distance problem, for a long time people tried to use various elementary methods to prove better lower bounds. There was a major breakthrough by Guth–Katz about ten years ago:

Theorem 4.9 (Guth–Katz 2015)
n points in $\mathbb{R}^2$ form $\Omega(n/\log n)$ distinct distances.

Remark 4.10. This is not necessarily tight for the number of distinct distances, but it *is* tight for the square grid if you use a L^2 distribution count.

All the other results had n to some exponent less than 1; this gets the correct exponent. It uses a lot of nice ideas, including incidence geometry (which we'll see) and classical algebraic geometry.

Remark 4.11. There is a relationship between these problems, similar to that of independent sets and chromatic number — if we look at n points, then

$$(\#\text{distinct distances}) \cdot (\max \#\text{unit distances}) \geq \binom{n}{2}$$

(we can make any distance a unit distance, by choosing a scale appropriately — there are $\binom{n}{2}$ distances, and each distinct one occurs at most (max #unit distances) number of times). This means an upper bound on (max #unit distances) implies a lower bound on (#distinct distances), but not vice versa.

This was a bit of an interlude showing a geometric application of KST (and more generally, a result in extremal graph theory). We'll now continue our exploration of various values of $\mathrm{ex}(n, H)$.

Last time, we saw the Erdős–Stone–Simonovits theorem:

Theorem 4.12 (Erdős–Stone–Simonovits)
We have

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

So if we ignore terms of order less than n^2, then we know the complete answer — in particular, for non-bipartite graphs this gives the correct asymptotic.

§4.2 Odd cycles

Now let's think about odd cycles. For triangles, we have the exact answer

$$\operatorname{ex}(n, C_3) = \left\lfloor \frac{n^2}{4} \right\rfloor.$$

What about longer (odd) cycles? We know by ESS that

$$\operatorname{ex}(n, C_{2k+1}) = \left(\frac{1}{4} + o(1)\right) n^2.$$

But it turns out that you can say much more — for every integer k,

$$\operatorname{ex}(n, C_{2k+1}) = \left\lfloor \frac{n^2}{4} \right\rfloor$$

for all sufficiently large n (this is again achieved by the complete bipartite graph).

We won't prove it, but we'll mention it because it's a nice result, and maybe somewhat counterintuitive.

First, it's consistent with ESS — it's a much sharper version. Interestingly, if you replace C_{2k+1} by some 3-chromatic graph, you might not get the same conclusion — on the pset, we'll come up with a 3-chromatic graph where the RHS is much larger. So there is something about odd cycles that is important.

And lastly, it's important n is large — this is false for small values of n. (If n is really small, this cannot be true.)

Roughly, the proof idea is that from ESS, you get that you have at least $(1/4+o(1))n^2$. Now suppose you're very close to that many edges. Then there are some results about stability that say that if you're very close to this number of edges, then the structure of your graph must be approximately bipartite. Once you are down to an almost bipartite graph, you have a much better handle of what can go on; and then you basically have to optimize the number of edges using the complete bipartite graph.

(So you use ESS to get close to the correct bound, then stability.)

§4.3 Even cycles

Meanwhile, there's a lot more mystery for even cycles.

The smallest even cycle is C_4, i.e., a $K_{2,2}$. We know that

$$\operatorname{ex}(n, C_4) = \operatorname{ex}(n, C_{2,2}) = O(n^{3/2})$$

by KST. What about longer even cycles?

Theorem 4.13

For all $k \geq 2$, there exists a constant C so that

$$\mathrm{ex}(n, C_{2k}) \leq Cn^{1+1/k}.$$

When $k = 2$, we get $n^{3/2}$ as above; when $k = 3$ we get $n^{4/3}$; and so on.

Remark 4.14. First, is this tight? We will see tomorrow that it is tight (up to constant factors) for some specific values of k, namely 2, 3, and 5. For all the other values of k, it is a major open problem whether this bound can be improved. In particular, for 8-cycles, we have an upper bound of $n^{5/4}$, and we do not know if that is sharp — it's open whether $\mathrm{ex}(n, C_8) = \Theta(n^{5/4})$.

We won't prove this, but we'll show a weaker result:

Theorem 4.15

We have $\mathrm{ex}(n, \{C_4, C_6, \ldots, C_{2k}\}) = O_k(n^{1+1/k})$.

This notation refers to the maximum number of edges in a n-vertex graph that avoids *all* of C_4, C_6, …, C_{2k} — i.e., without any even cycles of length *at most* $2k$. (So we simultaneously forbid all these cycles from appearing. This is a weaker result compared to the original theorem, where we only forbid one cycle.)

We'll prove this theorem — if we have too many edges, then we must have a short even cycle. So you give me a graph with lots of edges, and I want to produce a short even cycle.

If we can find a cycle, there's a quick and dirty way to ensure that a cycle has even length — first do some cleaning of the graph to make it bipartite.

Lemma 4.16

Every graph G has a bipartite subgraph with at least half of the edges.

So we can keep at least half of the edges and delete the other half, making our graph bipartite.

Proof. This is a classic application of the probabilistic method — color each vertex red or blue, uniformly and independently at random, and keep only the edges with one red and one blue endpoint. Every edge has probability $\frac{1}{2}$ of being kept, so the expected number of edges that are kept is exactly the total number of edges divided by 2; therefore there must be some event (i.e., choice of vertex colors) where we have at least this many edges left. □

So by losing at most half of our edges, we can go down to a bipartite graph.

This is part of some general ideas in graph theory where we do cleaning steps to get to a graph that's much easier to work with.

In the next lemma, we'll show that starting with a graph with lots of edges (i.e., large average degree), by removing not too many edges, we can get to a graph with large *minimum* degree.

Lemma 4.17

Let $t > 0$. Every graph with average degree $2t$ has a subgraph with minimum degree greater than t.

We start with a graph with lots of edges; but there may be some vertices with too few edges. The idea is to just get rid of those vertices; but we have to be a bit careful with execution.

Proof. We claim that removing a vertex of degree at most t cannot decrease the average degree. To see this, we start with average degree $2e(G)/v(G)$. Then we remove one vertex, and at most t edges; so our average degree is now at least

$$\frac{2e(G) - 2t}{v(G) - 1} \geq \frac{2e(G)}{v(G)}$$

(this is because $2e(G)/v(G) \geq 2t$).

So if we start with some average degree and remove a vertex of degree at most t, then our new average degree is at least the old one. So we can keep doing this — we ask, is there any vertex with degree at most t? And if there is, we get rid of it, and keep going. Eventually there will be no vertices of degree less than t, and you're done. (Note that we can't end up with the empty graph, because before that we'd have had the one-vertex graph, which doesn't satisfy this condition.) □

That's our second cleaning step; now we'll combine these lemmas to prove the theorem.

Proof. Suppose that G has no even cycles of length at most $2k$. We're going to first apply the two cleaning steps. We first apply the first lemma to get a bipartite subgraph G' with average degree at least $2t$, where $t = e(G)/2v(G)$, and then apply the second lemma to get a subgraph $G'' \subseteq G'$ with minimum degree greater than t. (Up to constant factors, we haven't changed anything — but now our graph is both bipartite, and has large minimum degree.)

We now do a breadth-first search. All the vertices we're left with have large degree. So we pick an arbitrary vertex, which has degree at least t going out. Each one of its neighbors also has degree at least $t-1$ outwards (since it has one edge back), and so on. Furthermore, if we call these layers A_1, A_2, …, A_k, two vertices in A_2 can't have a common neighbor in A_3, since then we could trace out a cycle.

More precisely, let A_i be the set of vertices at distance exactly i from v. Because of bipartiteness, each of these sets must be an independent set (otherwise we'd violate bipartiteness — an edge inside A_i would give us an odd cycle). Furthermore, each vertex in A_i has exactly one neighbor in A_{i-1} — if it had two neighbors backwards, then we could trace back to form an even cycle, which would violate the even cycle-free condition. (This is true all the way up to A_k.)

So every vertex has at least $t - 1$ outgoing edges to the next layer (since we have degree at least t, and only one edge going back). Putting everything together, the total number of vertices in the graph is at least the number of vertices in the final layer, which is at least $(t-1)^k$. So we have

$$v(G) \geq |A_k| \geq \left(\frac{e(G)}{2v(G)} - 1\right)^k.$$

Then rearranging this calculation gives the desired bound

$$e(G) \leq 2v(G)^{1+1/k} + 2v(G).$$ □

So this proof shows that if we forbid *all* the even cycles of length up to $2k$, then we have this upper bound. The idea is to first clean up the graph by making it bipartite and with large min degree. Then if we start with a large vertex, if we take k steps out from it, there can't be any coalescing. So the kth layer has to be really large; but it can't be larger than the number of vertices, which gives us an upper bound (on the number of edges).

> **Remark 4.18.** Why does this not prove the stronger result — e.g. that $C_6 = O(n^{4/3})$? If you only forbid C_6, we might have some C_4's, and then the proof would not work. Roughly speaking, the way to get around this is that if you only forbid C_6, then even if you see a C_4, you can sort of grow it into C_6.

§4.4 Bounded maximum degree

We'll continue looking at upper bounds for forbidding bipartite graphs; we'll now try to forbid a bipartite graph with bounded maximum degree.

We saw that if we have a bipartite graph H, then KST gives some bound on its extremal number. That's not always very good; we'll prove a theorem that's sometimes better.

Theorem 4.19

Let H be a bipartite graph with vertex bipartition $A \cup B$, such that every vertex in A has degree at most r. Then there exists a constant $C = C_H$ such that for all n, we have

$$\operatorname{ex}(n, H) \le Cn^{2-1/r}.$$

How does this compare to KST? KST would give us that

$$\operatorname{ex}(n, H) = O_H(n^{2-1/|B|}).$$

(You'd first say that H is a subgraph of a complete bipartite graph, and then apply KST on that complete bipartite graph; that gives you a worse bound.)

This bound in particular generalizes KST, because you can also apply it when H is a complete bipartite graph (in which case it gives the same number). Because of that, in this formulation the exponent cannot be improved; however, for specific graphs H one might be able to improve it.

Example 4.20

When $H = C_6$, if we apply KST, we get that $\operatorname{ex}(n, C_6) \le \operatorname{ex}(n, K_{3,3}) \lesssim n^{5/3}$. Meanwhile this theorem gives a better bound of $n^{3/2}$, which coincides with the bound for C_4's. However,that's not the truth — the tight bound is $n^{4/3}$ (which is what we discussed, though we didn't prove it).

Besides being interesting on its own, one reason to show us this theorem is that the proof technique is very interesting. It uses an important method in the probabilistic method, known as *dependent random choice*.

§4.4.1 Dependent Random Choice

Here's the setup: we're given a graph G, and G has many edges (for some sense of 'many'); we want to find an embedding of our graph H into G.

The goal of dependent random choice is to find a large subset $U \subseteq V(G)$ such that every r-vertex subset of U has many common neighbors in G. Think of r as e.g. 2; we want to find a large subset of vertices so that there's a lot of commonality, in that whenever we pick a pair of vertices in U, they have lots of common neighbors.

Intuitively, how might we achieve this? We mentioned the probabilistic method; one naive attempt is to try to pick U randomly (keeping every vertex with some probability). This is not going to work very well — a priori there's no reason it'd help us get lots of common neighbors.

So what else can we do? Imagine a social analogy — let's say you want to form a party, and invite some people to this party. You want to make sure that to be an interesting and fun party, every pair of participants have lots of common friends (so they haev something to talk about). If you invite people uniformly at random, you likely will not get that. But what you can do is pick someone who is very social, and invite all of *their* friends. If you invite all of a specific person's friends, then it's certainly more likely that they'll pairwise have more common acquaintances — they have at least this one person, and maybe others too.

If you want to do this even more, you can try picking a few people, and inviting their common friends altogether. That more or less works, except that you might not get exactly this property — maybe some of the invitees will just know those people and no one else. There won't be too many such people though, and you can just disinvite them.

We'll now write them down more precisely. The statement has a bunch of parameters, which we won't worry about because they just come from the proof.

Theorem 4.21 (Dependent random choice)

Let n, r, m, t be positive integers, and let $\alpha > 0$. Then every graph G with n vertices and at least $\alpha n^2/2$ edges contains a vertex subset U with

$$|U| \geq n\alpha^t - \binom{n}{r}\left(\frac{m}{n}\right)^t$$

such that every r-element subset S of U has more than m common neighbors in G.

Proof. We say that an r-element subset of $V(G)$ is *bad* if it has at most m common neighbors in G.

In this proof, some things will be random and some won't; this is somewhat confusing, so we'll write the random quantities in a different color. Let $u_1, \ldots, u_t$ be vertices chosen independently and uniformly at random from $v(G)$, allowing repetitions. Let A be the common neighborhood of $u_1, \ldots, u_t$. (These are the promoters in our social analogy, and we invite their common friends.)

For each (fixed) vertex $v \in V(G)$, what's the probability it's adjacent to all of $u_1, \ldots, u_t$? Well, v has some neighborhood; if v ends up adjacent to all these vertices, each one must land in its neighborhood. Each does independently with $(\deg v)/n$ probability, and there are t such vertices, so v has probability $(\deg v/n)^t$. This means by linearity of expectation that

$$\mathbb{E}[|A|] = \sum_{v \in V(G)} \mathbb{P}(v \in A) = \sum_{v \in V(G)} \left(\frac{\deg v}{n}\right)^t.$$

By convexity, we can lower-bound this by $n\alpha^t$ (the average of the numbers $\deg v/n$ is α by our assumption).

The idea is to choose A; A is a pretty large set, and hopefully it has nice properties. But it doesn't quite work yet — there might be some bad sets in A.

For any $R \subseteq V(G)$, we have

$$\mathbb{P}(R \subseteq A) = \mathbb{P}(R \text{ is complete to } u_1, \ldots, u_t).$$

Since these are independent uniform vertices, the probability this occurs is the number of common neighbors of R divided by n, raised to the tth power.

So if R is bad, then we have very few common neighbors (namely, at most m), and then

$$\mathbb{P}(R \subseteq A) \leq \left(\frac{m}{n}\right)^t.$$

Summing over all $\binom{n}{r}$ possible r-element subsets, we get

$$\mathbb{E}[\#\text{bad } r\text{-vertex subsets of } A] \leq \binom{n}{r} \cdot \left(\frac{m}{n}\right)^t.$$

Finally, we're going to obtain U by starting with A, and deleting a vertex from each bad r-vertex subset — we start with A, and if we see some bad subset, we delete an element arbitrarily. We do this for all the bad subsets, and in this way we destroy all of them. Then

$$\mathbb{E}\,|U| \geq \mathbb{E}\,|A| - \mathbb{E}[\#\text{bad } r\text{-vertex subsets of } A] \geq n\alpha^t - \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

So there exists U of at least this size satisfying the desired properties. □

Now let's use the dependent random choice theorem to prove the extremal theorem we stated earlier.

Proof. Let G be a n-vertex graph with lots of edges — more precisely, at least $Cn^{2-1/r}$ edges. We'll show that if C is large enough, we can embed H into G.

By choosing C to be large, we can make sure that the expression in the dependent choice theorem with $t = r$ is reasonable — α is $2Cn^{-1/r}$, so we can ensure that

$$n(2Cn^{-1/r})^r - \binom{n}{r}\left(\frac{|A|+|B|}{n}\right)^r \geq |B|\,.$$

(the first term, the two r's cancel out and by choosing C to be large we can make it a large constant; meanwhile A and B are constants, and both things have degree r in n, so the second term is a constant not depending on C). Then by DRC with $t = r$, we can embed B as U in the dependent random choice lemma, into G (we view B as the vertices of H; we embed H having A and B into our graph G, so B goes somehwere), such that every r-vertex subset of B (viewed as a subset of $V(G)$) has more than $|A| + |B|$ common neighbors.

The dependent random choice lemma tells us how to embed B into this graph G; what's next? Look at some vertex $a \in A$; it has at most r neighbors in A. We want to ask, can I embed this vertex into G? Well, if theh neighbors of A are some r things, then we look at where these land; and we want to make sure there's some choice for their common neighbor. And those vertices do share a common neighbor because of dependent random choice — in fact, they share lots of common neighbors. You need lots of common neighbors because maybe some of the points were used earlier in the process (taken up by other vertices), but there's still lots of vertices left — so each vertex in A can be embedded (there's lots of room left to embed A, even if we've embedded all the other vertices — we can always pick one of the common neighbors that is still unused). That finishes the proof, since the theorem says that if C is large we can embed a copy of H into G, and we've done so. □

> **Remark 4.22.** Does this let you count embeddings? If you apply DRC as a black box to get the existence of a single B, then you don't get very far. But we actually have that not just B exists, but you can get it with some probability; and then you can get lots of choices.

This wraps up our discussion of upper bounds. Next lecture we'll talk about lower bounds (for bipartite graphs).

§5 September 20, 2023

The last few lectures, we've been proving various upper bounds on the extremal number — what's the maximum number of edges in a n-vertex graph avoiding a specific graph H?

Now we'll turn our attention to lower bounds. We saw one sort of lower bound when $\chi(H) \geq 3$ — taking the Turán graph gives a lower bound of quadratic order. Today we'll look at what happens when H is bipartite (where there's a lot more mystery), and we'll see various ways to construct good H-free graphs with lots of edges.

§5.1 Randomized constructions

Here's the idea: we want to construct a H-free graph with lots of edges. The basic idea is to take a random graph, specifically $G(n,p)$ (the Erdős–Rényi random graph where we have n vertices and each edge occurs with probability p). If p is small enough, this graph will be H-free with high probability. But you can do something slightly better — if p is small, then likely $G(n,p)$ only has *few* copies of H. So we can make this graph H-free by deleting an edge from each copy (this destroys every copy of H).

What does having a *few* copies of H mean? We want the number of copies to be much less than the number of edges — then we can destroy all the copies of H without substantially changing the number of edges. The number of edges is $\Theta(n^2p)$, so we want $o(n^2p)$ copies of H.

Theorem 5.1

Let H be a graph with at least 2 edges. Then

$$\mathrm{ex}(n,H) \gtrsim_H n^{2-\frac{v(H)-2}{e(H)-1}}.$$

(The notation $\gtrsim_H$ means up to a constant factor depending on H.) This gives a H-free graph with lots of edges.

Proof. We'll follow the outline mentioned earlier. We first pick G to be a copy of $G(n,p)$, with p chosen as

$$p = \frac{1}{4} n^{\frac{-v(H)-2}{e(H)-1}}$$

(you can choose this value by working backwards). Then $\mathbb{E}[e(G)] = p\binom{n}{2}$. Let X denote the number of copies of H (as a subgraph) in G. We can count $\mathbb{E}[X]$ very precisely because it's some combinatorial expression, but we only care about the asymptotics, so we'll do this crudely — there are at most $n^{v(H)}$ ways to place the copy of H, and each occurs with probability $p^{e(H)}$, so

$$\mathbb{E}[X] \le p^{e(H)} n^{v(H)}.$$

Comparing this with $\mathbb{E}[e(G)]$ for our specific value of p, we get $\mathbb{E}[X] \le \frac{1}{2}\mathbb{E}[e(G)]$ (our value of p was chosen so that this particular inequality is true).

Then we have

$$\mathbb{E}[e(G) - X] \ge \frac{1}{2} \cdot p\binom{n}{2} \gtrsim n^{2-\frac{v(H)-2}{e(H)-1}}.$$

(Note that we get $v(H)-2$ and $e(H)-1$ because we compared $p^{e(H)}n^{v(H)}$ to pn^2, so we subtract 2 from v and 1 from e).

So there exists a graph G for which $e(G) - X$ is at least this expectation, and then taking this graph G and removing one edge from every copy of H yields a H-free graph with at least the desired number of edges. □

In a second, we'll do some comparisons to see how good this bound is. But first we'll comment that for certain H's, we can do slightly better. For example, when H is the graph obtained by starting with a $K_{4,4}$ and adding on two vertices, each of which we connect to two (different) vertices on the right-hand side (so $v(H) = 10$ and $e(H) = 20$), we get $\mathrm{ex}(n,H) \gtrsim n^{2-8/19}$. But we can do better. The point is that in order to avoid H, it's enough to avoid $K_{4,4}$; and that's actually 'harder' to avoid, so we'll do that instead. Then $K_{4,4}$ has 8 vertices and 16 edges, so this theorem gives us $\mathrm{ex}(n,K_{4,4}) \gtrsim n^{2-6/15}$, which is bigger than the earlier number; the conclusion is that

$$\mathrm{ex}(n,H) \ge \mathrm{ex}(n,K_{4,4}) \gtrsim n^{2-6/15},$$

which is better than applying the theorem to H directly.

So the lesson from this example is that instead of applying the theorem to H itself, we should look at all subgraphs of H and see which one produces the best result.

For this, we'll define a notion known in the literature as 2-density:

Definition 5.2. The 2-density of a graph H is

$$m_2(H) = \max_{H' \subseteq H, e(H') \geq 2} \frac{e(H') - 1}{v(H') - 2}.$$

Corollary 5.3

We have $\mathrm{ex}(n, H) \gtrsim_H n^{2-1/m_2(H)}$.

To prove this, we essentially apply the theorem to all subgraphs of H and use the one that gives the best exponent in the asymptotic.

Let's do some specific numerical comparisons. For $K_{2,2}$, what have we learned? From KST, we had an upper bound that tells us that

$$\mathrm{ex}(n, K_{2,2}) \lesssim n^{3/2}.$$

Meanwhile, plugging in $K_{2,2}$ into this theorem gives

$$\mathrm{ex}(n, K_{2,2}) \gtrsim n^{4/3}.$$

When you see situations like this, you want to ask, which one is the truth (or is the truth in between)? We'll later see that the upper bound is the truth — the random lower-bound construction can be improved.

More generally for $K_{s,t}$ with $t \leq s$, KST and this theorem together give

$$n^{2-\frac{s+t-2}{st-1}} \lesssim \mathrm{ex}(n, K_{s,t}) \lesssim n^{2-1/s}.$$

For any specific values of s and t there is a gap between these, but they do get closer and closer for large values of s.

That's all we'll say for randomized constructions. The nice thing about them is that this construction works for all graphs H; but it might not be very good (it's believed to more or less be never tight — you can probably come up with some silly cases where it might be tight, but we don't know of any interesting ones where it is tight).

Remark 5.4. There was a recent breakthrough on the Kahn–Kalai conjecture, about whether there exists a copy of H in $G(n,p)$. For that, you can calculate the threshold for having H. To find this threshold, you might want to look at a subgraph of H. This corresponds to 1-density instead of 2-density.

So the randomized construction is general, but at least for this problem, it is rarely tight. In the rest of this lecture, we'll see a different way of constructing graphs based on algebra. We'll see several examples of tight constructions. They're very elegant, but more ad hoc — each graph requires a new idea, and we often don't know how to generalize these ideas to other graphs.

§5.2 Tight bound for $K_{2,2}$

Theorem 5.5

We have $\mathrm{ex}(n, K_{2,2}) \geq (\frac{1}{2} - o(1))n^{3/2}$.

We won't worry too much about the constant in front, but in fact it's the same constant as in the upper bound. In other words, from KST we also deduced that $\mathrm{ex}(n, K_{2,2}) \leq (\frac{1}{2} + o(1))n^{3/2}$.

The construction will be algebraic; before giving it, it's helpful to see some geometric intuition (because what's happening is really geometry). The geometric intuition is that the graph we're constructing is essentially what's known as a *point-line incidence graph*. We will be looking at some geometry (in the very general sense of the word, where we have points and lines). We'll have a bipartite graph where the left part corresponds to points, and the right part corresponds to lines. A point p and a line ℓ will be adjacent if and only if the point p lies on the line ℓ in this geometry. So we have points and lines, and we draw edges if the corresponding point lies on the corresponding line.

Claim 5.6 — Any point-line incidence graph is C_4-free.

(Note that $C_4 = K_{2,2}$.)

Proof. A C_4 would correspond to two points p_1 and p_2, and two lines ℓ_1 and ℓ_2, such that p_1 and p_2 both lie on ℓ_1 and ℓ_2. This is not possible, because every two points have at most one line going through them (in the kinds of geometries we'll look at). □

It turns out that if you try to do this in real Euclidean space, it will not work (in the sense it will not give you a good bound); this has to do with the topology of the real plane. Instead, we do this construction in a finite field geometry — we'll be looking at the finite field plane $\mathbb{F}_p^2$, where there are p^2 points and $p^2 + p$ lines. (You could consider the projective plane, but it doesn't make much of a difference.) Then using this construction, we see that

$$\mathrm{ex}(2p^2 + p, K_{2,2}) \geq p^3$$

(each of the at least p^2 lines contains p points each, so we have at least p^3 edges). This proves the result, at least when n is of a certain form (namely, having to do with primes or prime powers).

To extend to all n, we can invoke results from number theory, which say that the largest prime below n is close to n:

Theorem 5.7

The largest prime below N is $N - o(N)$.

You might know the classical result due to Chebyshev that there's always a prime between N and $2N$; this is good enough if you don't care about the constant factor. To get the right constant you need this result, which follows from the prime number theorem — if there are linear-sized gaps, then this violates the density given by the prime number theorem.

Remark 5.8. The state of the art result in this area is due to Baker–Harmon–Pintz (2001), which states that there exists a prime in $[N - N^{0.525}, N]$ for all large N.

This more or elss proves what we want, but if you work out the expressions, you don't actually get the constant $\frac{1}{2}$. Now we'll see how to get the constant $\frac{1}{2}$.

To do this, the idea is to treat the points and the lines as the same set of vertices. (There's a notion in projective geometry called point-line duality, where there's a duality between points and lines, and they play dual roles.) This is the idea of the *polarity graph*.

Proof. We let p be the largest prime such that $p^2 - 1 \le n$ (so that $p = (1 - o(1))\sqrt{n}$).

Let G be the graph with vertex set $V = \mathbb{F}_p^2 \setminus \{(0,0)\}$, with an edge $(x,y) \sim (a,b)$ if and only if $ax + by = 1$ in $\mathbb{F}_p$.

Here we're using pairs of $\mathbb{F}_p$-elements to denote both points and lines — here (a,b) describes the equation of a line and (x,y) a point, and the point (x,y) lies on the line corresponding to (a,b) if and only if this equation holds.

Note that any distinct (a,b) and (a',b') in V have at most one common neighbor — this is precisely because we're talking about two lines, and two lines intersect in at most one point (by solving this system of equations).

Then everything works the same as before — this graph is $K_{2,2}$-free. To find the number of edges, there are $p^2 - 1$ vertices, and each vertex has degree at least $p - 1$; then

$$\#\text{edges} \ge \frac{1}{2}(p^2-1)(p-1) = \left(\frac{1}{2} - o(1)\right) n^{3/2}. \qquad \square$$

§5.3 Construction for $K_{3,3}$

Theorem 5.9

We have $\mathrm{ex}(n, K_{3,3}) \ge (\frac{1}{2} - o(1))n^{5/3}$.

This again agrees with KST, up to a constant factor. (We won't worry about the leading factor; the exponent is what matters.)

Of course a $K_{2,2}$-free graph is also $K_{3,3}$-free, but the one we constructed doesn't have enough edges; so we need new ideas. The underlying geometric idea there was that two lines intersect in at most one point. In this case, we'll consider unit spheres. Two unit spheres intersect in a circle (of radius at most 1), and that circle will cross another unit sphere in at most two points; so three unit spheres have at most 2 common points (in $\mathbb{R}^3$) — this is the key geometric idea.

As before, we won't be able to work in real Euclidean space; we'll work over finite fields. Unfortunately, the proof we just described is a proof in Euclidean space (by drawing pictures of spheres). But it is actually an algebraic proof as well. In pictures, we said that two spheres intersect in a common circle, and that circle intersects another sphere in at most two points. But if you weren't allowed to handwave and draw pictures, you could prove this algebraically by writing down certain expressions; that would end up using the fact that a quadratic has at most two roots. And that algebraic proof more or less works over other fields as well.

Proof sketch. As before, let p be the largest prime such that $p < n^{1/3}$ (so it's not much less than $n^{1/3}$). Fix a nonzero $d \in \mathbb{F}_p$, which we take to be a quadratic residue mod p (i.e., the square of a nonzero element) if $p \equiv 3 \pmod 4$, and a quadratic nonresidue if $p \equiv 1 \pmod 4$. (We'll see why soon.)

We'll construct a graph G where $V = \mathbb{F}_p^3$, and we draw an edge $(x,y,z) \sim (a,b,c)$ if and only if

$$(a-x)^2 + (b-y)^2 + (c-z)^2 = d.$$

(In Euclidean space, this would say that the two points have a certain prescribed distance; over $\mathbb{F}_p$ there isn't a distance notion, so it's just some quadratic equation.)

This is a complete description of the graph; we now need to check a few things.

First we need to check that it has enough edges. It turns out that every vertex has lots of neighbors — more precisely, $(1 - o(1))p^2$ neighbors. Why? Here's some intuition: imagine we fix a, b, and c, and ask

how many solutions there are for x, y, and z. Here x, y, and z vary over the entire space of triples of $\mathbb{F}_p$ elements. For now imagine that they're chosen uniformly at random — so then we have a sum of 3 i.i.d. quadratic residues. The point is that if we sum 3 uniform quadratic residues together, as a random variable you'd expect the sum to be close to uniform. This requires a proof, which we're not going to do now; in fact it is somewhat related to things we'll see later in the course when we do Fourier analysis. (The best way to prove this is to invoke Fourier estimates, which we won't worry about for now.) So if they're uniform, the probability we hit a specific d should be $\frac{1}{p}$.

Remark 5.10. What would have happened if we instead randomized d, and took the 'best' one? It turns out that we can't do this, because there's another step in which we'll need the quadratic residue/nonresidue condition.

Remark 5.11. It's worth knowing that taking two i.i.d. quadratic residues is not enough.

It remains to show that the graph is $K_{3,3}$-free. For this part, we're going to handwave somewhat. How you show $K_{3,3}$-freeness in the Euclidean setting would involve some system of equations. So we can imagine first computing the equation for the radical plane between two spheres (the plane at which they intersect). Then we have three radical planes, and we want to take their intersection.

So we're essentially doing algebraic manipulations to mimic what happens in Euclidean space. The thing that we end up needing is that there's no sphere with 3 collinear points. (The intersection of the radical planes is a line, called the radical axis; that cuts through all the circles, and that's where the common intersections are. In Euclidean space you use the fact that the line intersects each circle in at most 2 points. That's not always true in $\mathbb{F}_p$, but with the condition on d, we can make it true.)

Putting everything together, this proves the desired result. □

Remark 5.12. What's the intuition for the quadratic (non)residue definition? We have a line and a sphere, and you can think about solving the equation of their intersection. For a bad choice of d you might actually have a sphere containing a line; but for appropriate d this doesn't happen.

We've gotten $K_{2,2}$ and $K_{3,3}$; what about $K_{4,4}$? It turns out that this is open. We have the KST upper bound of $n^{7/4}$.

Conjecture 5.13 (Open) — The KST bound on $K_{4,4}$ is tight — $\operatorname{ex}(n, K_{4,4}) = \Theta(n^{7/4})$.

No one knows how to generalize what happens here to deal with $K_{4,4}$. More generally, for $s \le t$ we have $\operatorname{ex}(n, K_{s,t}) \lesssim n^{2-1/s}$, but we don't know if this is tight. (Later we'll see how to prove something like this when s is *much* less than t.)

Remark 5.14. Announcements: PS1 is due Sunday, to be submitted on Gradescope. We should read all the instructions for submission — we should make sure we start each solution on a different page (to streamline grading) and include sources and collaborators on every problem (not just the first page). The submission deadline is 11:59pm; we should not wait to submit at 11:59pm (due to technical issues). We should start early. There are OH on Fridays; we should feel free to come chat about anything related to the class, especially the homework.

§5.4 The case $t \gg s$

We'll keep talking about $K_{s,t}$-free constructions. Earlier we saw $K_{2,2}$ and $K_{3,3}$ constructions that were tight. Next, we'll show that KST is tight when t is much larger than s.

Theorem 5.15

If $t > (s-1)!$ then $\mathrm{ex}(n, K_{s,t}) = \Theta_{s,t}(n^{1-2/s})$.

(This actually generalizes the previous results, up to the constant factors.)

First, we'll show this for $t > s!$.

Proof. As before, let p be a prime. We define the *norm map* $\mathcal{N}(x)$ in the following way — if we start with a field extension $\mathbb{F}_{p^s}$, then there's a norm map $\mathcal{N}\colon \mathbb{F}_{p^s} \to \mathbb{F}_p$ given by multiplying x by all its Galois conjugates. Explicitly, we set

$$\mathcal{N}(x) = x \cdot x^p \cdot x^{p^2} \cdots x^{p^{s-1}} = x^{\frac{p^s-1}{p-1}}.$$

This is some map, and the point is that even if we start with an element of the field extension, we get an element of $\mathbb{F}_p$ — one way to check this is to note that $\mathcal{N}(x)^p = \mathcal{N}(x)$ (and the set of points invariant under raising to the pth power is precisely $\mathbb{F}_p$).

Furthermore, since $\mathbb{F}_{p^s}^\times$ is the cyclic group of order $p^s - 1$ (this is also a fact from finite fields), we know that the number of elements $x \in \mathbb{F}_{p^s}$ with $\mathcal{N}(x) = 1$ is exactly $\frac{p^s-1}{p-1}$ (since in the cyclic group, we're just doing multiplication by this number).

The construction we'll use is known as the *norm graph* — we define the *norm graph* as the graph with vertices $V = \mathbb{F}_{p^s}$ and an edge between distinct elements $a, b \in \mathbb{F}_{p^s}$ if $\mathcal{N}(a+b) = 1$. (So there's some algebraic equation that describees the relationships between vertices that form edges, similarly to earlier.)

We can compute various quantities, like the minimum degree — given a, there are $\frac{p^s-1}{p-1}$ elements b for which adding that element would give us norm 1, but one of these may be a itself (which should be excluded); so every vertex has degree at least $\frac{p^s-1}{p-1} - 1$. And since there are p^s vertices, the number of edges is at least $\frac{1}{2}p^{2s-1}$. (This is the correct asymptotic order for the number of edges; we don't care about lower-order terms.)

The harder part is to show that this norm graph indeed is $K_{s,t}$-free for $t = s! + 1$ (which will mean we've found a $K_{s,t}$-free graph with enough edges). This is an algebraic fact, analogous to things we saw earlier (two lines intersect in at most one point, three spheres intersect in at most two). The proof relies on the following result (which was proved specifically for this application, but is nice on its own):

Theorem 5.16

Let $\mathbb{F}$ be any field, and let a_{ij} and b_i be elements of $\mathbb{F}$ such that $a_{ij} \neq a_{kj}$ for any $i \neq k$. Then the system of equations

$$\begin{aligned}(x_1 - a_{11})(x_2 - a_{12}) \cdots (x_s - a_{1s}) &= b_1\\ (x_1 - a_{21})(x_2 - a_{22}) \cdots (x_s - a_{2s}) &= b_2\\ &\vdots\end{aligned}$$

has at most $s!$ solutions $(x_1, \ldots, x_s) \in \mathbb{F}^s$.

To understand what's going on, it's helpful to think about the special case where $b = 0$. Then in the first row, we need to pick one of the factors to vanish; let's say we pick $x_1 = a_{11}$. The hypothesis tells us that down each column there are no repetitions; so once we've done this none of the other terms in the first column vanish, and we need to do something else for the remaining rows. SO in this case, there are exactly $s!$ solutions.

The hard part is what happens when $b \neq 0$. We won't say more; the proof is about a page of commutative algebra and algebraic geometry (Prof. Zhao has read it and can follow it line by line, but doesn't understand what's going on).

But once we have this theorem, we are more or less done — to see that the norm graph is $K_{s,t}$-free for $t = s! + 1$, consider distinct $y_1, \ldots, y_s \in \mathbb{F}_{p^s}$; we want to bound the number of common neighbors of x. A common neighbor is a solution to the system $1 = \mathcal{N}(x + y_i)$ (where we want to solve the system in x). This expands to

$$1 = (x + y_i)(x + y_i)^p \cdots (x + y_i^{p^{s-1}}).$$

From basic facts about finite fields, we know that $(x + y)^p = x^p + y^p$ (everything else consists of binomial coefficients divisible by p), so we can expand this as

$$1 = (x + y_i)(x^p + y_i^p) \cdots (x^{p^{s-1}} + y_i^{p^{s-1}}).$$

We run through $1 \leq i \leq s$, giving a system of s equations; and these satisfy the hypothesis of the theorem (we can check that the hypothesis $a_{ij} \neq a_{kj}$ is satisfied, because raising to the pth power is a bijection in this field). Then we can apply the theorem to get that there are at most $s!$ solutions, which then implies that the graph is $K_{s,t}$-free. □

We'll now talk about how we can go from the weaker result ($t = s! + 1$) to the stronger result ($t > (s-1)!$). To get the wider range of t, we modify the construction to the *projective norm graph* — here the vertex set will be *pairs* of elements with one in the field extension and the other in the multiplicative group of the field, i.e., $V = \mathbb{F}_{p^{s-1}} \times \mathbb{F}_p^\times$, where $(X, x) \sim (Y, y)$ if and only if $\mathcal{N}(X + Y) = xy$. Then this graph turns out to have just the right properties to let us deduce the claim; it turns out that you still apply the same theorem (there's no new ideas).

§5.5 Avoiding cycles

Now we'll talk about cycles; specifically, we'll show tight lower bounds for C_4, C_6, and C_{10}. Previously, we saw an upper bound for even cycles:

Theorem 5.17

We have $\operatorname{ex}(n, C_{2k}) = O_k(n^{1+1/k})$.

(We proved a weaker result.) We'll show that this bound is tight for three specific cycles; it's open for all $k \notin \{2, 3, 5\}$ (specifically, it's open for C_8).

In fact, there will be a unified proof that shows all of these case simultaneously, though that's not how it was discovered. We've already seen C_4 — there we did a complete point-line incidence graph, where we took all the points and all the lines in the plane. Here we'll still do a point-line incidence graph, but we'll restrict to certain points and lines.

Proof. Let q be a prime (or prime power). Let $\mathcal{L}$ be the set of all lines in $\mathbb{F}_q^k$ whose direction can be written as $(1, t, t^2, \ldots, t^{k-1})$ for some t. (This is sometimes known as the *moment curve* — you pick all directions of this form, and look at the corresponding lines. Not every line can be written in this way — when $k = 2$, all non-vertical lines can be written in this way, but for larger k it only represents a small subset of lines.)

Our construction is to look at the bipartite point-line incidence graph where we have points in $\mathbb{F}_q^k$ on the left-hand side, and $\mathcal{L}$ on the right-hand side. As before, we have an edge between p and ℓ if the point p lies on the line ℓ.

First we need to check that there are enough edges. This is a fairly straightforward calculation — every line has exactly q points, and there are q^k lines, so the number of edges is q^{k+1}, and the number of vertices is $2q^k$. So we do have enough edges.

The more interesting thing we want to show is that this construction avoids the cycle of length $2k$.

Proposition 5.18

For $k = 2$, 3, or 5, this construction is C_{2k}-free.

What does such a cycle look like? Well, a cycle would correspond to a sequence of alternating points and lines $p_1, \ell_1, \dots, p_k, \ell_k$ — this means p_1 and p_2 lie on ℓ_1, p_2 and p_3 on ℓ_2, and so on. So a $2k$-cycle corresponds to a k-gon in this geometry.

The line going from p_i to p_{i+1} is in our set, so we need to have

$$p_{i+1} - p_i = a_i(1, t_i, \dots, t_i^{k-1})$$

for each $i = 1, \dots, k$ (with indices mod k). But we're also going all the way around; so these must sum to 0. This gives us an equation

$$0 = \sum (p_{i+1} - p_i) = \sum a_i(1, t_i, \dots, t_i^{k-1}).$$

There are various directions that occur here; some may repeat (in the sense that some may be parallel), and we can delete duplicates. Then these directions (after deleting duplicates) are linearly independent — this is because of the Vandermonde determinant, since we know that

$$\begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{k-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{k-1} \\ \vdots & \vdots & \ddots & \vdots & \end{vmatrix} = \prod_{i<j} (x_i - x_j),$$

so if all the x's (which we can think of as our t's) are distinct, then this determinant is nonzero and these directions are linearly independent.

But we're taking a linear combination of them that sum to 0. How does this happen? This means every single one of them must cancel itself out — so every direction must appear at least twice.

But also, all the lines we're using are distinct, so no direction can come up twice in a row. So we have various directions, each must come up at least twice, and none can occur twice in a row. This cannot happen for two directions, three directions, or five directions, which finishes the proof. □

Remark 5.19. This doesn't work for an 8-cycle because we could have one direction happening on opposite sides of the 4-gon, and the other happening on the other opposite sides. In fact, whether the upper bound is tight for the 8-cycle is still a major open problem. Similarly, you can assign colors to a 7-gon such that each is used at least twice with no two in a row, and so on; so this doesn't work for any other values of k.

Remark 5.20. There is one section in the chapter that we won't cover, but that is beatiful — there are recent developments that combine these into randomized algebraic constructions (where instead of taking a specific polynomial, you take a random one). This does very well — a recent paper by Bukh improved the dependence $t > (s-1)!$ to roughly $t > 9^s$.

§6 September 25, 2023

We'll spend the next few lectures talking about the graph regularity method. This is a powerful method in extremal combinatorics. It's also not easy if we've never seen this kind of combinatorics before, because it looks kind of different from the rest of combinatorics. So we'll spend today talking about what Szemerédi's regularity lemma is (the statement and proof); and in subsequent lectures we'll see how to apply it to prove nice combinatorial facts.

Szemerédi's graph regularity lemma informally says the following:

Theorem 6.1 (Szemerédi's graph regularity lemma, informally)

The vertex set of every large graph can be partitioned into a bounded number of parts so that the graph looks random-like between most pairs of parts.

This is a very informal statement; in particular, what does 'random-like' mean? That's the next thing we'll explain.

You can think of the regularity lemma, at a very high level, as a structure vs. randomness (or signal vs. noise) decomposition. Given an arbitrary graph, no matter how large, there's a way to study it by considering a partition of the vertex set so that there's only a bounded number of parts (that's structure) and between parts everything looks random (that's pseudorandomness). This is very powerful, and the theorem was proven by Szemerédi in the 1970s. (He's an important figure in the field, and this is one of his contributions that had lasting impact in combinatorics.)

§6.1 Random-like graphs

First we'll define what we mean by random-like.

Definition 6.2. The edge density between $X, Y \subseteq V(G)$, denoted $d_G(X,Y)$, is

$$d_G(X,Y) = \frac{e_G(X,Y)}{|X| \cdot |Y|},$$

where we define $e_G(X,Y)$ as the number of *ordered* pairs $(x,y) \in X \times Y$ such that $xy \in E(G)$.

Note that X and Y are allowed to overlap, and this definition still makes sense. (For example, if $X = Y$ then $e_G(X,X)$ is twice the number of edges in X, and we're dividing by n^2 — equivalently, this gives the number of edges divided by $n^2/2$.)

The notion of random-like we'll use is captured in the following definition of an ε-regular pair.

Definition 6.3. Let G be a graph, and $U, W \subseteq V(G)$. We call (U,W) an *ε-regular pair* in G if for all subsets $A \subseteq U$ and $B \subseteq W$ with $|A| \geq \varepsilon |U|$ and $|B| \geq \varepsilon |W|$, we have

$$|d_G(A,B) - d_G(U,W)| \leq \varepsilon.$$

This is saying that — suppose we have two vertex sets U and W, and think about the bipartite graph between them. We want to say that it's rather homogeneous, in that even if we restrict down to small subsets (which are not *too* small), the edge density is similar to that of the original graph. (U and W don't have to be disjoint, but it's easier to visualize this when they are.)

Definition 6.4. If (U,W) is *not* ε-regular, then we say that their irregularity is *witnessed* by $A \subseteq U$ and $B \subseteq W$ if (A,B) if (A,B) is a counterexample to the previous definition.

There are three ε's that appear in the definition; the last plays a different role. You can refine the definition by isolating them, but it doesn't matter; we'll use the same one for convenience.

The reason that we call this 'random-like' is that if we generate a random graph between U and W, then with high probability it will satisfy this condition.

That's what it means to be random-like *between* a pair of parts. Now, what does it mean to have a partition? For that, we need the following definition:

Definition 6.5. Given a graph G, a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ (where $V_1, \dots, V_k$ are subsets of $V(G)$) is an *ε-regular partition* if

$$\sum |V_i|\,|V_j| \le \varepsilon\,|V(G)|^2$$

where the sum is over all ordered pairs $(i,j) \in [k]^2$ where (V_i, V_j) is *not* ε-regular.

In other words, at most an ε-fraction of pairs of vertices lie between irregular pairs.

Remark 6.6. It's helpful to think about the special case when all the vertex sets in this partition have the same size. In this case, the condition is equivalent to having at most εk^2 irregular pairs — you have k^2 pairs of parts, and if all parts have the same size, then the condition says that of all these k^2 pairs, at most an ε-fraction of them are irregular.

It's important that we allow *some* irregular pairs — what we want to prove turns out to be false if we want every pair to be regular. So this is why we instead require that if we look at the irregular pairs, there aren't too many — but there may be some exceptions.

§6.2 Szemerédi's graph regularity lemma

Now we're ready to state the main theorem.

Theorem 6.7 (Szemerédi's graph regularity lemma)

For every ε, there exists M depending only on ε such that every graph has an ε-regular partition into at most M parts.

So if you give me some error tolerance ε (an arbitrarily small constant), I can produce for you another constant M (which only depends on ε, and not the graph) so that no matter what graph you give me, no matter how large, I can partition the vertex set into at most M parts so that the graph is ε-regular in the above sense. This is a precise formulation of the informal version that we stated earlier.

The goal of the rest of today is to prove this regularity lemma.

Remark 6.8. Why do we say 'large' in the informal version? If M is 1000 and we only have a 100-vertex graph, then we can trivially get a ε-regular partition by putting every vertex in its own part. So for small n, the statement is true but not useful. (This subject is different from classical graph theory where you play with 5-vertex graphs and see what happens.)

Remark 6.9. Why did we need A and B to be large in the definition of ε-regular pairs? If we drop this, then this is too strong of a condition to ask for — for example, if we take A and B to be a single vertex, then the edge density between them fluctuates between 0 and 1. So we certainly need some lower bound on the sizes of A and B — otherwise this is too strong.

Remark 6.10. In the next chapter, we'll discuss pseudorandomness much more deeply.

§6.3 Proof Sketch

We'll first see the idea of the proof. We'll come up with an algorithm to generate this partition. The algorithm will start with a trivial partition, with exactly one part — i.e., $\mathcal{P} = \{V(G)\}$.

While this partition is not ε-regular (so we're not done yet), we're going to identify all the pairs of parts that are ε-irregular (i.e., not ε-regular), and find a witnessing pair.

We're then going to simultaneously refine this partition $\mathcal{P}$ into a new partition $\mathcal{Q}$ by introducing all these pairs — a partition is a way to cut up the vertex set, and each time we see an irregular pair, we find a further refinement to cut up the parts even further (and we cut them up along these witness pairs).

The while loop will stop when the partition becomes regular. *A priori*, it might go on for a long time; so we want to bound the number of steps, so we can bound the number of parts we end up with.

We will bound the number of steps using an *energy increment argument* — we'll keep track of something we call the energy of the partition, so that at each step the energy has to go up by a definite amount. Since the energy will always be bounded between 0 and 1, this process cannot continue too long.

So that's the proof strategy; now we'll execute it.

§6.4 Energy

We'll define the notion of energy in a few steps. Here G is a n-vertex graph; we will drop the dependence on G, so G stays the same throughout this argument.

Definition 6.11. For all $U, W \subseteq V(G)$, define

$$q(U,W) = \frac{|U| \cdot |W|}{n^2} \cdot d(U,W)^2.$$

Definition 6.12. For every partition $\mathcal{P}_U = \{U_1, \dots, U_k\}$ of U and $\mathcal{P}_W = \{W_1, \dots, W_\ell\}$ of W, we define

$$q(\mathcal{P}_U, \mathcal{P}_W) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} q(U_i, W_j).$$

Definition 6.13. For a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$, we define its energy as

$$q(\mathcal{P}) = q(\mathcal{P}, \mathcal{P}).$$

In other words,

$$q(G) = \sum_{i,j=1}^{k} q(V_i, V_j) = \sum_{i,j=1}^{k} \frac{|V_i|\,|V_j|}{n^2} d(V_i, V_j)^2.$$

So this is some kind of weighted square mean of edge densities. That's also why this is called an energy — it's a L^2-quantity, and many quantities in physics that are energies are L^2.

This is the definition; now we'll prove a number of properties about energy.

Fact 6.14 — We always have $0 \le q(\mathcal{P}) \le 1$.

This is because it's a weighted sum of edge densities (squared), which are always between 0 and 1.

Lemma 6.15

Energy never decreases under refinement — if $U, W \subseteq V(G)$ and $\mathcal{P}_U$ is a partition of U and $\mathcal{P}_W$ is a partition of W, then

$$q(\mathcal{P}_U, \mathcal{P}_W) \geq q(U, W).$$

So if we start with U and W and we break them up, the energy cannot go down — it can only go up.

Proof. Imagine that we have two sets U and W, and we break them up according to $\mathcal{P}_U$ and $\mathcal{P}_W$. We want to consider the energies between them.

This basically follows from convexity, but to write this out more explicitly: let $n = V(G)$ (we'll use this throughout), and let $\mathcal{P}_U = \{U_1, \dots, U_k\}$ and $\mathcal{P}_W = \{W_1, \dots, W_\ell\}$. Choose $x \in U$ and $y \in W$ uniformly and independently, and suppose that they fall into the parts U_i and W_j. Define the random variable Z to be $Z = d(U_i, W_j)$. (In other words, we're picking two random vertices x and y, and looking at the edge density of the two containing parts — this is a random variable because we don't know which parts we're choosing.) Then

$$\mathbb{E}[Z] = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{|U_i|}{|U|} \cdot \frac{|W_j|}{|W|} \cdot d(U_i, W_j)$$

(we consider all the different possibilities, with their corresponding probabilities). But this is precisely

$$d(U, W) = \sqrt{\frac{n^2}{|U| \cdot |W|} q(U, W)}.$$

On the other hand, we have

$$\mathbb{E}[Z^2] = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{|U_i|}{|U|} \cdot \frac{|W_j|}{|W|} \cdot d(U_i, W_j)^2 = \frac{n^2}{|U|\,|W|} q(\mathcal{P}_U, \mathcal{P}_W)$$

(we basically defined q as the weighted square density, but with different normalization).

But we have $\mathbb{E}[Z^2] \geq (\mathbb{E}[Z])^2$ (by convexity), which gives exactly what we want. □

So that's our first lemma — energy never decreases under refinement. We'll now look at another lemma, which says something in similar spirit.

Lemma 6.16

Given two partitions $\mathcal{P}$ and $\mathcal{P}'$ of $V(G)$, if $\mathcal{P}'$ refines $\mathcal{P}$, then $q(\mathcal{P}) \leq q(\mathcal{P}')$.

So if we can obtain $\mathcal{P}'$ by starting at $\mathcal{P}$ and cutting up more parts, then the refined partition has at least as much energy as the coarse one.

Proof. The proof is by summing the previous lemma over all pairs of parts — we start with some partition $\mathcal{P}$, and then go to some refinement $\mathcal{P}'$ by cutting up parts. We want to check that tne energy of $\mathcal{P}'$ is at least that of $\mathcal{P}$. We can do this part-by-part, for example by comparing the energy between two parts in $\mathcal{P}$ with the energy between what we cut them into. We can do this for all pairs, including for a part against itself, and summing gives the desired result. □

So far, we've said that if we refine, the energy can never go down. That's not enough — maybe it never goes down, but never goes up much either. So what we next want to show is that if we start with something irregular, then the energy must increase substantially when we refine it.

Lemma 6.17 (Energy boost for irregular pairs)
If (U, W) is not ε-regular, as witnessed by $A \subseteq U$ and $B \subseteq W$, then we have

$$q(\{A, U \setminus A\}, \{B, W \setminus B\}) > q(U, W) + \varepsilon^4 \frac{|U|\,|W|}{n^4}.$$

So we start with irregular (U, W), and we refine them by cutting them into two parts — we cut U into A and its complement in U, and W into B and its complement. We know that this must increase the energy (weakly), but in fact it must increase by some given amount.

Remark 6.18. Prof. Zhao likes to call this the red bull lemma — if you're feeling irregular, it gives you a boost in energy.

Proof. Use the same setup as before, where we pick a random variable $x \in U$ and $y \in W$, and define Z to be the edge density between their respective parts.

Last time we only used the fact that $\mathbb{E}[Z^2] \geq (\mathbb{E}[Z])^2$, but in fact the difference between these two is their variance — we have

$$\operatorname{Var} Z = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \frac{n^2}{|U|\,|W|}(q(\mathcal{P}_U, \mathcal{P}_W) - q(U, W)).$$

We want to show that this quantity is bounded below by some quantity. To show that a variance is bounded below, we just need to show it can take on two very different values with reasonable probability. And that's essentially the definition of being irregular — there are two very different possibilities that can appear.

So we have $Z = d(A, B)$ with probability at least

$$\frac{|A| \cdot |B|}{|U| \cdot |W|}$$

(this tracks the probability that x and y are in A and B, respectively). So just looking at this event, we have

$$\operatorname{Var} Z = \mathbb{E}[(Z - \mathbb{E}Z)^2] \geq \frac{|A| \cdot |B|}{|U| \cdot |W|}(d(A, B) - d(U, W))^2.$$

In the definition of a witnessing pair, each of the ratios $|A|/|U|$ and $|B|/|W|$ is at least ε, and the quantity inside the square is greater than ε in absolte value. So this entire expression is greater than ε^4, and putting this together with our expression for $\operatorname{Var}(Z)$ in terms of q gives the lemma. □

§6.5 Proof of Regularity Lemma

Here's where we're at — we want to prove Szemerédi's regularity lemma, and we're trying to implement the proof strategy where we start with a trivial partition, and keep refining it according to its irregularities. We want some way to show this process stops after a finite number of steps (that's a way to control the finite number of parts). And we saw that a useful way to keep track of our progress is this notion of energy, which is sort of a squared mean of densities between parts. We showed that refinement can never decrease the energy; and if we start with a single pair and break it up according to its witness in irregularity, the energy must go up by some definite amount. The next step is to put these things together, and think about what happens when we start with an irregular partition.

Lemma 6.19

If a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$ is not ε-regular, then there exists a partition $\mathcal{Q}$ that refines $\mathcal{P}$ where every part V_i is partitioned into at most 2^{k+1} parts, such that

$$q(\mathcal{Q}) > q(\mathcal{P}) + \varepsilon^5.$$

This is the crucial step — here we're showing that if we start with an irregular partition, then we can refine it further so that the energy goes up by at least ε^5 (which is a constant). We're going to basically implement the strategy described above, but there's some trickiness to it.

Proof. Let $R = \{(i,j) \in [k]^2 \mid (V_i, V_j) \text{ is } \varepsilon\text{-regular}\}$ (here we allow i to equal j), adn let $\overline{R} = [k]^2 \setminus R$ be the set of iregular pairs.

For each $(i,j) \in \overline{R}$, we can find a witnessing pair $A^{ij} \subseteq V_i$ and $B^{ij} \subseteq V_j$.

So if we have three parts, we look at V_1 and V_2, and ask if they're irregular. If so, we can find some witnessing pair for V_1 and V_2. Similarly, between V_1 and V_3 we find some witnessing pair; and between V_2 and V_3 we find some witnessing pair (for now, suppose that they're irregular with each other). Some parts might also be irregular with themselves; suppose that this happens, so we can also find witnessing pairs for those parts.

We then simultaneously put all these individual sets together into a common refinement — we cut up our three parts even further by introducing *all* the cuts along these witnessing pairs. This will be our resulting partition (where we introduce all of these cuts simultaneously).

There are some redundancies — for example, we have $A^{ij} = B^{ji}$ due to symmetry for all $i \neq j$.

Set $\mathcal{Q}$ to be the common refinement of $\mathcal{P}$ and all of these A^{ij} and B^{ij} (where we put all of them together). How many parts are there? Each V_i is refined by at most $k+1$ subsets — $k-1$ from the other V_j's, and possibly 2 from within V_i itself. This gives at most 2^{k+1} parts in refinement (which is where the 2^{k+1} comes from).

This bounds the number of parts in the refinement; now we need to show that the energy of $\mathcal{Q}$ is substantially larger than that of $\mathcal{P}$. We'll do this invoking the earlier lemmas — define $\mathcal{Q}_i$ to be the partition of V_i given by $\mathcal{Q}$ (so e.g. $\mathcal{Q}_2$ is the partition of V_2). Then we can break up $q(\mathcal{Q})$ as

$$q(\mathcal{Q}) = \sum_{(i,j) \in [k^2]} q(\mathcal{Q}_i, \mathcal{Q}_j) = \sum_{(i,j) \in R} q(\mathcal{Q}_i, \mathcal{Q}_j) + \sum_{(i,j) \in \overline{R}} q(\mathcal{Q}_i, \mathcal{Q}_j).$$

For the regular pairs, we just apply the fact that the energy doesn't decrease — so we have

$$\sum_{(i,j) \in R} q(\mathcal{Q}_i, \mathcal{Q}_j) \geq \sum_{(i,j) \in R} q(V_i, V_j).$$

Meanwhile, for the irregular pairs, for each $(i,j) \in \overline{\mathcal{R}}$ we had a very complicated partition, but by monotonicity we're allowed to forget some of the cuts — we can just look at the partition corresponding to the witness. So we have

$$\sum_{(i,j) \in \overline{R}} q(Q_i, Q_j) \geq \sum_{(i,j) \in \overline{R}} q(\{A^{ij}, V_i \setminus A^{ij}\}, \{B^{ij}, V_j \setminus B^{ij}\}) > \sum_{(i,j) \in \overline{R}} \left(q(V_i, V_j) + \frac{\varepsilon^4 |V_i| |V_j|}{n^2} \right)$$

(using the energy-boosting lemma). Now let's look at the second term; because of the definition of an ε-regular pair, this quantity is strictly bigger than ε^5 (failing to have ε-regularity makes this sum big). Together, we get the lemma that we're trying to prove. □

We presented this proof in a somewhat careful but deliberate way. There's a reason we do this: if you just follow the idea we outlined, you might do something else — suppose we look at a partition that currently has three parts, and see that it's not regular. Then we might find an irregular pair, and break it up according to that irregular pair. And then we look at these new parts, and if that's not irregular we break it up according to some witness; and so on. This is in a way also a natural stretegy, but it doesn't work — because you cannot guarantee a sufficiently large energy increment. The point of this proof is that we're identifying *simultaneously* all the irregularity, and then simultaneously refining the partition (not one at a time); that's how we get the result. It's a subtle difference, but an important one.

We now have all the main ingredients, and we're almost there.

Proof of SRL. Start with the trivial partition $\mathcal{P} = \{V(G)\}$, and repeatedly apply the previous lemma whenever the partition $\mathcal{P}$ is not ε-regular. Because the energy is always between 0 and 1, we must stop after at most ε^{-5} steps. (So the energy increment bounds the number of steps in this process.) Each step takes us from a partition with k parts to another partition with at most $k \cdot 2^{k+1}$ parts. So the number of parts is bounded as a function of ε.

It turns out to be convenient to upper bound this quantity by 2^{2^k}; then the number of parts in the end is at most $2^{2^{2^{\cdots}}}$ where the tower has height at most $2\varepsilon^{-5}$. (Each step increases our tower height by at most 2, and we have at most ε^{-5} steps.) So we can define M as this quantity (which depends only on ε). □

§6.6 Some remarks

There are many obvious questions we should ask, one of which is — what's up with this bound? It's a very large number, and you might ask, is it large because we did this proof in a suboptimal way? Initially people thought that might be the case, but it turns out you really need this many parts:

Theorem 6.20 (Gowers)

For all $\varepsilon > 0$, there exists a graph with no ε-regular partition having fewer than $2^{2^{\cdots}}$ parts where the tower has height ε^{-c} for a small constant c.

So apart from the exponent of -5, the bound given by this proof really is optimal (we can't do better).

This was quite surprising when found, because we're going to be applying the regularity lemma to various applications, and when we do that we have to incur the very poor quantitative bounds that come with it.

There are a few more comments to make. One is that in the definition of an ε-regular partition, we had to allow for some number of irregular pairs — we cannot guarantee that all pairs are regular. That turns out to also be important — asking for all pairs to be regular is too strong, and there will be a homework problem exploring this fact. It gives us a specific counterexample:

Example 6.21

The *half graph* is a bipartite graph where $i \sim j$ if and only if $i \le j$ (each vertex is connected to all the vertices horizontal or below).

It turns out that we need some irregular pairs in order to partition this graph.

Another comment is that the version presented here allows parts to have different sizes. That's what comes out of the proof, but it's not a serious requirement — you can actually guarantee all the parts have basically the same sizes.

Theorem 6.22 (Equitable regularity lemma)

For all $\varepsilon > 0$ and m_0, there exists a constant M such that every graph has an ε-regular equitable partition into k parts for some $m_0 \le k \le M$, where *equitable* means that all part sizes are within 1 of each other.

Here we have a lower bound m_0 on the number of parts. We can ignore this because it's convenient to ignore parts against themselves, and if all the parts are small enough then they don't matter. So this is mostly as a matter of convenience, and it doesn't affect things too much.

When we proved SRL, we repeatedly refined and increased the energy, but this didn't give us control over the sizes of the vertex sets. How do we get this control? This class is much more like analysis than algebra — we can ignore ε's everywhere — and the details are somewhat annoying, but they can be done.

One thing we can do is: we get a partition at the end that might not be equitable, and we're going to try to cut things up to make it equitable. This is hard because if you start with an ε-regular partition and refine it, it might not stay ε-regular — refinement might destroy ε-regularity.

Remark 6.23. In the proof we performed, we kept doing refinements. A misunderstanding of the proof is that each step makes the partition more and more regular. This is *not* what happens — we have no guarantee that after each operation we've done something to regularize the graph. All we're saying is that there's a magical quantity called energy that keeps the number of steps from being too large; but it doesn't tell us anything else.

So if we get a partition in the end and naively cut it up, we might destroy regularity. You can maybe get around this with randomness, but an easier way is to enforce equitability at every step. Then when you refine things energy might go down a *little* bit because you have to move things around to keep things regular, but if you gain ε^{-5} and lose half of that, you still gain energy. So you can prove the theorem in this way.

Another funny thing is that for every ε, there exists a partition with at most some number of parts. Do we know how many parts it has? The proof doesn't actually tell us a specific number, it just gives an upper bound — and it's not like we can always attain exactly that number of parts. (It says we do this and eventually something is good, but it doesn't give us a guarantee on a specific number of parts.) So we should not write that we have an ε-regular partition into exactly M parts — that is *not* guaranteed by the theorem.

In PS3, we'll be given quite a few problems that ask us to apply the regularity lemma. It's sometimes more convenient (as a matter of notation) to apply the equitable version, and we are allowed to do that (though we'll use the version we proved in the next few lectures).

This subject can be a bit of a learning curve because there's a lot of stuff to keep track of. It's important not to get bogged down too much by the individual δ's and ε's — most applications might have lots of error parameters in their statements, and it's important not to get too focused on these specific parameters and instead focus on the ideas.

§7 September 27, 2023

Today we'll continue our discussion of the regularity method.

Last time, we stated and proved the Szemerédi regularity lemma. Roughly speaking, it says that given a very large graph, we can partition it to a bounded number of pieces so that the graph looks random-like between most pairs of parts — between one pair it might look like a random graph with density 0.1, between another pair it might look random-like with density 0.4, and so on.

Today we want to see the other part of the regularity method, and use it to prove some applications. The goal of today's lecture is to give a proof of Roth's theorem, that a subset of $[N]$ with no 3-APs has very few elements.

§7.1 Triangle counting lemma

In order to use the regularity method, we need some additional tools. One of these tools is called a *counting lemma*.

What can we do with the pseudorandom decomposition? We'll prove something called the counting lemma, which roughly speaking, says the following: suppose you have three vertex sets X, Y, and Z (think of these as parts in our regularity partition), and we want to know how many triangles there are among X, Y, and Z. If we had a random graph with densities α, β, and γ, we'd expect the number of triangles to be roughly $|X|\,|Y|\,|Z|\cdot\alpha\beta\gamma$. The counting lemma says that in our setting, this is roughly true — the number of triangles is roughly what is expected from looking at edge densities.

Now let's formulate this more precisely.

Lemma 7.1 (Triangle counting lemma)
Let G be a graph, and let X, Y, and Z be three vertex subsets of G such that (X,Y), (Y,Z), and (X,Z) are all ε-regular and have edge density at least 2ε. Then the number of triangles $xyz \in X \times Y \times Z$ is at least

$$\#\{xyz \in X \times Y \times Z \text{ triangles}\} \geq (1-2\varepsilon)(d(X,Y)-\varepsilon)(d(Y,Z)-\varepsilon)(d(X,Z)-\varepsilon)\cdot |X|\,|Y|\,|Z|\,.$$

So up to some small ε error tolerances, the claim says that the number of triangles xyz is roughly what you should expect.

It's fine for now to think of X, Y, and Z as disjoint sets; but the lemma also allows them to overlap (similarly to how we defined the number of edges between X and Y). In practice we'll only use the lemma when X, Y, and Z are either disjoint or identical (we can think of X, Y, and Z as various parts).

Remark 7.2. Is there also an upper bound? We'll just prove the lower bound today, but a similar proof also gives an upper bound; later when we talk about graph limits, we'll see another proof that simultaneously gives both, but the proof we'll see here can easily be adapted.

Remark 7.3. We shouldn't worry too much about the specific forms of these expressions, like in analysis — what matters is the methods.

To prove this, we'll first prove a lemma that intuitively says that most vertices have roughly the same degree in a regular pair (at least, coming from one side).

Lemma 7.4
Suppose that (X,Y) is an ε-regular pair. Then fewer than $\varepsilon\,|X|$ vertices in X have fewer than $(d(X,Y)-\varepsilon)\,|Y|$ neighbors in Y.

So we have two vertex sets X and Y; a priori they are ε-regular. This lemma says that the degrees can't vary too much — we don't have too many vertices in X with very small degree to Y.

This is mostly an exercise in understanding the definition of an ε-regular pair — it's a statement about uniformity and it doesn't say anything about any specific vertex, but it does say you can't have too many exceptional vertices.

Proof. Let A be the subset of X containing all the vertices with fewer than $(d(X,Y)-\varepsilon)|Y|$ neighbors in Y (i.e., the undesirable vertices). Then there are very few edges between A and Y, because there are very few out-degrees coming from A — in particular, we have

$$d(A,Y) < d(X,Y) - \varepsilon,$$

because every vertex in A has density less than $d(X,Y)-\varepsilon$. But because of the definition of regularity, this shouldn't happen unless A is too small — if A is large, then this violates the definition of regularity. So since (X,Y) is ε-regular, we must have $|A| < \varepsilon|X|$; this is the desired conclusion. □

Now let's use this several times to prove the triangle counting lemma. The rough idea is that we start with three vertex sets. A typical vertex in X (up to some small exceptions) has lots of neighbors in Y and lots of neighbors in Z. These neighborhoods are all pretty large; so then we can apply regularity between Y and Z to see that there must be lots of edges between these two neighborhoods, which gives you lots of triangles.

Now let's work this out more carefully.

Proof of triangle counting lemma. By the previous lemma, by excluding at most a small fraction of vertices (applying the lemma to edges outgoing to Y as well as to Z), we can find a set $X' \subseteq X$ with $|X'| \geq (1-2\varepsilon)|X|$ such that every vertex $x \in X'$ has lots of neighbors to both Y and Z — more precisely, at least $(d(X,Y)-\varepsilon)|Y|$ neighbors in Y and at least $(d(X,Z)-\varepsilon)|Z|$ neighbors in Z. We write $N_Y(x) = N(x) \cap Y$ and $N_Z(x) = N(x) \cap Z$.

For all $x \in X'$, we have $|N_Y(x)| \geq (d(X,Y)-\varepsilon)|Y| \geq \varepsilon|Y|$, and likewise $|N_Z(x)| \geq \varepsilon|Z|$. So we have two large neighborhoods; and then we can apply the definition of a regular pair once more to generate lots of edges between these two neighborhoods — we have

$$\#\{\text{edges between } N_Y(x) \text{ and } N_Z(x)\} \geq (d(Y,Z)-\varepsilon)|N_Y(x)||N_Z(x)|,$$

using the definition of regularity again. Now plugging in all the inequalities gives the originally claimed bound. (The first factor of $(1-2\varepsilon)|X|$ comes from the number of choices for the first vertex; then $(d(X,Y)-\varepsilon)|Y|$ and $(d(X,Z)-\varepsilon)|Z|$ describe the neighborhood sizes, and $d(Y,Z)-\varepsilon$ comes from the number of edges between them.) □

§7.2 Triangle removal lemma

So we now have two tools — the regularity lemma giving us a partition, and the counting lemma giving us one way to use the information coming from the partition. Now we'll put these together to prove an important result, known as the triangle removal lemma.

Lemma 7.5

For every ε, there exists δ such that every n-vertex graph with fewer than δn^3 triangles can be made triangle-free by removing at most εn^2 edges.

This statement says that if we have a graph with very few triangles (i.e., the number of triangles should be on the order of n^3), then we can delete a small number of edges to make the graph triangle-free.

Remark 7.6. This is not just supersaturation — we haven't yet seen anything like this, and it's much harder in a way than what we've seen in the first chapter. It *looks* simple — if you have few triangles, you can get rid of them by killing a few edges. But which edges should you take? Even though there's few triangles, few means sub-cubic, but you only want to delete a sub-quadratic number of edges. So this looks deceptively simple, but it's actually very difficult.

We can also write the statement in a different, equivalent form:

Lemma 7.7

An n-vertex graph with $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges.

We have to interpret the second statement correctly — there's an $o(n^3)$ in the hypothesis; what does this mean? One way to interpret it is by considering a *sequence* of graphs instead of a single graph — if we have a sequence of graphs with sub-cubic numbers of triangles, we can make them triangle-free by removing sub-quadratic numbers of edges.

This will be our first application of the graph regularity method; historically it was also one of the first applications of the regularity method. And also, in some sense, we don't know how to prove this any other way — it's completely different from anything we've seen before.

This is also a great way to explain how to apply the regularity method. We'll explain this using a representative recipe — most standard applications of the regularity method have proofs following three steps.

(1) Apply the regularity lemma to *partition* the vertex set.

(2) *Clean* the graph — we get a partition that has lots of nice properties, but there's also some exceptions (we allow some small number of irregular pairs). So it's not completely clean everywhere, and we have to do some work to clean the graph. Typically this means removing edges between irregular pairs, edges between low-density pairs, and when we're not using the equitable version of the regularity lemma, also removing edges from pairs involving small parts. (So we get a regularity partition, and do some work to get rid of edges that don't behave well.)

(3) Apply the counting lemma to show we can embed our structure (or triangle) into the cleaned up version of the regularity partition.

So this is the general recipe; let's see how to apply it to prove the triangle removal lemma.

Proof of triangle removal lemma. We're given a n-vertex graph G with few (less than δn^3) triangles; we'll choose δ at the end (it'll come out of the proof).

We apply Szemerédi's regularity lemma; this gives us an $\varepsilon/4$-regular partition of $V(G)$ into parts $V_1, \ldots, V_m$, where m is at most some constant that only depends on ε (we can write this as $m = O_\varepsilon(1)$).

That's the partition step (to apply the regularity lemma). Now we'll perform the cleaning step — for each pair of vertex parts $(i, j) \in [m]^2$ (allowing $i = j$), we remove all edges between V_i and V_j if any of the following hold:

- (V_i, V_j) is not $\varepsilon/4$-regular.
- $d(V_i, V_j) < \varepsilon/2$.
- $\min(|V_i|, |V_j|) < \varepsilon n/4m$.

So we get rid of irregular pairs (since they don't work well with what we're about to do); we also get rid of pairs where the density is too small (the reason is that in the triangle counting lemma we require the edge densities to be somewhat large, so if they're too small the lemma breaks). Finally, we get rid of edges between two parts where one of the vertex sets is very small. The *typical* vertex set size is n/m; so if it's quite a bit smaller than that, we say it's way too small and decide not to worry about these edges.

We should check that the number of edges removed is not too large. The number of edges removed in the first step is at most $\sum_{(i,j)} |V_i|\,|V_j|$ where the sum is over all irregular pairs (using a very crude upper bound for the number of edges between them). But the definition of a regular partition says that this quantity is not too large; in particular, it's at most $\varepsilon n^2/4$.

For the second type of operation, all the densities we remove are very small, so even if we do this *everywhere* we're not removing too many edges — we can perform this computation and check that we're removing at most $\varepsilon n^2/2$ edges. More carefully, the number of edges removed is

$$\sum d(V_i, V_j)\,|V_i|\,|V_j|$$

where the sum only consists of terms where $d(V_i, V_j) < \varepsilon/2$. Replacing it, this is at most

$$\sum \frac{\varepsilon}{2} |V_i| |V_j| \le \frac{\varepsilon}{2} n^2.$$

And in the third, we're getting rid of edges from parts with at least one small vertex set. There are m parts, and we're removing at most $\varepsilon n/4m \cdot n$ edges from each — for each part that is small, the part has at most $\varepsilon n/4m$ vertices, and even if it's connected to everything else it has at most $\varepsilon n/4m \cdot n$ edges. So this gives us an upper bound of $\varepsilon n^2/4$.

So all together, we've removed fewer than εn^2 edges. (This is what we're trying to do — remove a small number of edges.)

Our goal was to show that we can make the graph triangle-free, so let's hope that we've done that (provided that δ is chosen appropriately).

Claim 7.8 — The remaining graph is triangle-free.

Proof. Suppose that some triangle remains in the graph; and let's say that the triangle lies among three parts V_i, V_j, and V_k. (Here i, j, and k don't necessarily have to be distinct — there's just a triangle somewhere, and we look at the index at which each vertex lies.)

The idea is to use the triangle counting lemma to show that if we find such a configuration, then we get not just one triangle that remains, but actually a *ton* of triangles — on the order of cubically many (which will then contradict the hypothesis that the original graph had very few triangles).

We claim the three vertex sets V_i, V_j, and V_k satisfy the hypothesis of the triangle counting lemma — they're pairwise $\varepsilon/4$-regular, and have density at least $\varepsilon/2$, so we can apply the triangle counting lemma to obtain that the number of triangles in $V_i \times V_j \times V_k$ is at least

$$\left(1 - \frac{\varepsilon}{2}\right) \left(\frac{\varepsilon}{4}\right)^3 |V_i| |V_j| |V_k|.$$

And from our cleaning step we also know that none of these vertex sets are too small, so $|V_i| |V_j| |V_k| \ge (\varepsilon n/4m)^3$.

Recall that m is a constant — ε is given, and m only depends on ε. So as long as δ is a small enough constant (allowed to depend on ε but not on n), the number of triangles here is bigger than e.g. $6\delta n^3$. (The reason we write 6 here is that we allow the possibility $i = j = k$, in which case we overcount by a factor of 6; this is a constant factor, so we won't worry about it.)

So the number of triangles among these three parts is already larger than what we are allowed, which is a contradiction; so the remaining graph must atually be triangle-free. □

Let's recap the important steps of the proof. We follow the regularity recipe — first we partition the graph using Szemerédi's graph regularity lemma. This may have some imperfections — irregular pairs, low-density pairs, and small parts — and we get rid of them, which we can do without removing too many edges.

The claim is that now we've made the graph triangle-free. This is because if we didn't, we'd have some triangle; and this triangle must sit among three pairs that are pairwise ε-regular and have large density. And we apply the triangle counting lemma to then get lots and lots of triangles, contradicting the hypothesis. □

Remark 7.9. If there's $o(n^3)$ triangles, does this method give you a bound on the number of edges in the graph? The answer is no — a complete bipartite graph is triangle-free and has lots of edges. But the next thing we'll prove is close in spirit.

Corollary 7.10 (Diamond-free lemma)

If G is an n-vertex graph where every edge lies on a unique triangle, then the number of edges in G is $o(n^2)$.

The reason this is called a diamond-free lemma is that we're putting down lots of triangles in a way such that no two triangles share an edge; two triangles sharing an edge is sometimes called a diamond.

Proof. Suppose that G has m edges. Because every edge lies in a unique triangle, there must be exactly $m/3$ triangles (the edges are a disjoint union of triangles). But $m \le n^2$, so the number of triangles is $O(n^2)$, which in particular is $o(n^3)$.

Then by the triangle removal lemma, G can be made triangle-free by removing $o(n^2)$ edges. But there are $m/3$ triangles, and we need to remove an edge from each one of them; this menas we need to remove at least $m/3$ edges. So we must have $m = o(n^2)$. □

We mentioned this result in the first lecture; we mentioned that this is a deceptively simple statement (it looks like something you might be able to prove bare-hands with elementary graph theory, but that's not the cases). But we basically don't know any other proof.

Remark 7.11. We *do* know a small improvement, in the sense of the quantitative dependencies (how δ depends on ε, or what we put in the $o(n^2)$). The dependencies in the graph regularity lemma are tower-type, and we applied that lemma in our proof of the triangle removal lemma, so that comes into our value of m; here δ has to be smaller than the quantity $(\varepsilon n/4m)^3$, which involves m, so we need δ^{-1} to be a tower of height polynomial in ε^{-1}. And since we're applying that in the diamond-free lemma, $o(n^2)$ is n^2 over something that's essentially $\log_* n$, where $\log_*$ is the number of logs we need to apply to bring n to less than 1. (This is a very slowly growing function — it's the inverse of tower height.)

There has been one improvement, due to Jacob Fox — when he was a grad student, he found a proof where δ^{-1} is still a tower of exponentias, but of height $\log 1/\varepsilon$ rather than polynomial in $1/\varepsilon$. So the bound we get here is $n^2/e^{\log_* n}$. (This is still very slowly growing.) And that's the best that anyone knows — there hasn't been any improvement in either of these two theorems.

On the other hand, the best lower bound is very far off — next time we'll see the Behrend construction, which in particular implies thatin the diamond-free lemma, we can't do better than $n^2/e^{c\sqrt{\log n}}$. There's a huge gap between $\log_* n$ and $\sqrt{\log n}$, and that's the state of our knowledge. (It's basically the difference between something almost polynomial, and something that's a tower.)

Remark 7.12. The second and third problem sets have been released. We are strongly encouraged to start thinking about these regularity problems early, because it's one of the most conceptually difficult parts of the course.

§7.3 Roth's theorem

The next goal is to use the diamond-free lemma to prove Roth's theorem. This proof was not Roth's original proof; that uses Fourier analysis, and we'll see it later. This proof is from the 1970s by Rusza and Szemerédi, and it's one of the first applications of the regularity method.

Theorem 7.13

If $A \subseteq [N]$ is 3-AP-free, then $|A| = o(N)$.

In the first lecture of the course (in illustrating the connection between graph theory and additive combinatorics), we proved Schur's theorem by setting up a graph where solutions to $x + y = z$ correspond to triangles. We'll see a similar idea here — by setting up an appropriate graph, we'll be able to reduce Roth's theorem to the diamond-free lemma.

Proof. We start with the integers, but it's nicer to work in a group; so we embed A into $\mathbb{Z}/M\mathbb{Z}$ where $M = 2N + 1$. (The reason we increase N by a factor of 2 is to avoid wrap-arounds — we're embedding A inside part of a circle so that 3-APs in A over the integers correspond exactly to 3-APs in A in the cyclic group, since you can't jump around 0.)

We'll construct a tripartite graph G as follows. We'll have three vertex sets X, Y, and Z, which are all copies of the cyclic group $\mathbb{Z}/M\mathbb{Z}$. We draw edges according to various rules:

- We draw an edge xy for $x \in X$ and $y \in Y$ if and only if $y - x \in A$. (This is basically a bipartite Cayley graph.)
- Likewise, we draw an edge yz for $y \in Y$ and $z \in Z$ if and only if $z - y \in A$.
- Finally, we draw an edge xz for $x \in X$ and $z \in Z$ if and only if $\frac{z-x}{2} \in A$. (We can divide by 2 because we're in an odd cyclic group.)

So we've essentially made three bipartite Cayley graphs.

Why choose these rules? Note that $(x, y, z) \in X \times Y \times Z$ is a triangle if and only if $y - x$, $\frac{z-x}{2}$, and $z - y$ all lie in A; but the middle expression is the average of the other two, so they form a 3-AP.

But we started with a set A that was 3-AP free. Does that mean there are no triangles in this graph? Not quite — the caveat is that the 3-AP could be a constant 3-AP. But that's the only possibility — since A is 3-AP-free, these three numbers must be equal. In particular, if we're given x and y, this uniquely determines z (and likewise with the other pairs); this implies that every edge of G lies in a unique triangle, namely the triangle we get by solving the equality of these three numbers.

The number of vertices is roughly $6N$, so the diamond-free lemma says that G has $o(N^2)$ edges. But by construction we have $e(G) = 3M\,|A|$. Since M and N are on the same order, this implies that $|A| = o(N)$. □

Remark 7.14. Is it true that if $A \subseteq [N]$ is free of $x + y = z$, then $|A| = o(N)$? No — we can take the latter half, or just the odd numbers. What's the difference between avoiding sum-free and avoiding 3-APs? The point is that we need the constant 3-APs — there's such things as trivial 3-APs. This proof only could work if our pattern is 'translation-invariant' (if we take a 3-AP and shift everything by 100, then it's still a 3-AP; that's not true for sums).

§7.4 Some more results

Now let's prove a few more related theorems. Here we'll look at an instance of the multidimensional Szemerédi theorem, where we want to avoid 2-dimensional patterns. Here we'll try to avoid a corner, with two legs of the same length — explicitly, a corner is three points of the form (x, y), $(x, y + d)$, and $(x + d, y)$ for $d > 0$.

Theorem 7.15

Every corner-free subset of $[N]^2$ has size $o(N^2)$.

This was proved initially by Szemerédi, using a complicated proof that invoked the full Szemerédi theorem; later a short proof was found that only invokes the diamond-free lemma.

Proof. We'll essentially try to do something similar to before, but first we need to do a small trick to allow $d \neq 0$ starting from $d > 0$ (in our original definition a corner has a certain orientation; it shouldn't make much difference to allow negative d, but this needs proof). The point is that if $A \subseteq [N]^2$ is corner-free, then for each $z \in \mathbb{Z}^2$, we can let $A_z = A \cap (z - A)$ (where $z - A$ is a reflected version of A, translated appropriately). If we take a typical z, then this set should be large just by linearity of expectations — if we average over z, this should be pretty large, so there in particular exists one z for which this set is large. Explicitly, we have

$$\sum_{z \in [2N]^2} |A_z| = |A|^2,$$

so there exists some z such that

$$|A_z| \geq \frac{|A|^2}{(2N)^2}.$$

If we're trying to show that A has size $o(N^2)$, then it suffices to show that this specific A_z has size $o(N^2)$. And because of the symmetry introduced when defining A_z, we have that A_z is corner-free for all *nonzero* d, rather than just positive d (because we flipped A and intersected it).

The crux of the proof is the following: from now, we'll write A instead of A_z. Similar to before, we're going to build a tripartite graph G with three parts X, Y, and Z; and we need to put in edges between these parts.

The idea is to have each vertex $x_i \in X$ correspond to the vertical line at $x = i$, each vertex $y_j \in Y$ correspond to the horizontal line $y = j$, and each $z_k \in Z$ correspond to the -1-slope line $x + y = k$.

To decide whether to draw a line between x_i and y_j, we look at the horizontal line at i and the vertical line at j, and we put in an edge if and only if there's a point in A at their intersection — so $x_i \sim y_j$ if and only if $(i, j) \in A$.

Likewise, to determine if there's an edge between x_i and z_k, we look at the lines $x = i$ and $x + y = k$, and put in an edge if and only if we have a point in A at their intersection; and we do the same for edges between y_j and z_k (looking at the vertical line at j and -1-slope line at k).

Since A is corner-free, what are the triangles in this graph? If (x_i, y_j, z_k) form a triangle, then geometrically we have three lines which pairwise intersect in elements of A. This isn't possible because we don't have corners, *unless* these three points coalesce.

So then similarly to before, every edge of G lies in a unique triangle. (If we have an edge between x_i and y_j, then we have a horizontal and vertical line that intersect at one point; they form a triangle with the z_k corresponding to the diagonal line through that point.) And then the rest is the same, and we can again conclude that $|A| = o(N^2)$.

(Here $X = [N]$, $Y = [N]$, and $Z = [2N]$. Here we didn't need mods because we're looking at plane geometry.) □

Remark 7.16. What happens for higher-dimensional generalizations of corners? This is still true, but we need a hypergraph version of the triangle removal lemma, and for that we need a hypergraph regularity lemma (which is much more difficult to prove).

Remark 7.17. The corner-free and 3-AP-free problems are related to each other — let $r_3(N)$ be the size of the largest 3-AP-free subset of $[N]$, and $r_L(N)$ the size of the largest corner-free subset of $[N]^2$. Then we claim that

$$r_3(N) \cdot N \le r_L(2N).$$

The point is that given a 3-AP-free set $A \subseteq [N]$, we can construct a large corner-free set in the following way. We have a box of length $2N$, and some points of A on the x-axis from 0 to N. We then include all the points on the diagonal through these points of A, and take that to be our set B.

We see that $|B|$ is basically $|A| \cdot N$ (every line has slightly different size, but between N and $2N$ points). And if A is 3-AP-free then B is corner-free — if there were a corner in B, that corner projects down to a 3-AP in A.

The reason we mention this is to say that having an upper bound on corner-free sets gives another proof of Roth's theorem — the right-hand side being $o(N^2)$ implies Roth's theorem. These are basically the same proof — underlyingly we're using the diamond-free lemma, though the execution is somewhat different.

§8 October 2, 2023

The last couple of lectures, we've been talking about the regularity method. Last lecture, we used it to prove Roth's theorem — that every 3-AP-free subset of $[N]$ has $o(N)$ elements. First we proved Szemerédi's regularity lemma (about partitioning an arbitrary graph), and then we looked at the regularity recipe of how to apply it, as well as a counting lemma to prove a triangle-removal lemma that then implied Roth's theorem.

Today we'll begin by further discussing Roth's theorem.

Theorem 8.1 (Roth)
If $A \subseteq [N]$ is 3-AP-free, then $|A| = O(N)$.

The proof we gave last time gave a bound that was $o(N)$, but only barely (the quantitative bounds in Szemerédi's lemma are extremely poor) — it's roughly $N/\log_* N$. Later we'll see Roth's original proof using Fourier analysis, which will give much better bounds ($N/\log\log N$).

§8.1 A lower bound construction

Today we'll see a lower bound construction — how do you construct sets that are fairly large and still 3-AP-free?

One natural idea is to try building the set greedily — start with 0, and add numbers one at a time if they don't create a 3-AP. This gives 0, 1, 3, 4, 9, 10, 12, 13, 27, ... — this is known as the *Stanley sequence*. It's a fun exercise to figure out what this sequence is; the answer is that if you look at the sequence in base 3, then you get the sequence

$$0, 1, 10, 11, 100, 101, 110, 111, 1000, \ldots,$$

namely all the numbers whose base-3 representation has only 0's and 1's (and no 2's). It'll be left as an exercise to check that this is indeed the case.

How big is this set? When $N = 3^k$, then A has 2^k elements — so we can rewrite this as

$$|A| = 2^k = N^{\log_3 2}.$$

When N is not a power of 3, we get something slightly different, but not by much.

For a very long time, people thought this was the best you can do; for small N it's hard to beat this quantity.

So it was quite surprising when in the 1940s, Salem and Spencer found a set that was 3-AP-free set and had size *almost* linear.

Theorem 8.2 (Salem–Spencer 1942)
There exist $A \subseteq [N]$ which are 3-AP-free and have $|A| = N^{1-o(1)}$.

There was a subsequent improvement by Behrend (modifying some of these ideas):

Theorem 8.3 (Behrend 1946)
There exist 3-AP-free sets A with $|A| \geq Ne^{-C\sqrt{\log N}}$.

The Behrend construction is very important in additive combinatorics; to this date we haven't seen any substantial improvement over this construction (despite a lot of effort).

In a way, this might be the truth — two years ago it'd have been a mystery of whether we're closer to the Roth or Behrend bounds, but now we do know the answer, because of recent breakthroughs: we do know that every 3-AP-free set A has

$$|A| \leq Ne^{-(\log N)^{\beta}},$$

where β is some exponent (not necessarily $1/2$). So it's quite possible that the Behrend construction is close to the truth.

§8.2 The Behrend construction

The idea behind the Behrend construction comes up all over the place — it's one of the only ideas we have for constructing very large sets in additive combinatorics.

The idea is — I want a large set that avoids 3-APs. In the integers, this is hard to visualize — there are 3-APs all over the place. But there's a related geometric problem, that's a lot easier to visualize — on a sphere, you do not have 3-APs. In fact, there are no three points on a line in a sphere.

So what we can do is take a high-dimensional sphere and intersect it with some integer lattice; you can choose the sphere so that you get lots of points. (We'll do this by the pigeonhole principle.)

That's not quite a subset of the integers, but that's okay — we can then project this set down to the integers, using base-q expansion.

The point here is that because the initial set was free of 3-APs, this process doesn't change that process; so the subset of $\mathbb{Z}$ we get will also be free of 3-APs. And it turns out that if we choose the right parameters, this set will be fairly large.

So that's the strategy; now we'll go through the details.

Proof. We'll pick $m, d \in \mathbb{Z}$ later (for now, they're just some integers depending on N). Let $X = \{0, 1, \ldots, m-1\}^d$ be the d-dimensional box, and let

$$X_L = \{x \in X \mid |x|^2 = L\}.$$

As L ranges over all integers from 0 to dm^2, these spheres cover all possible points; so by pigeonhole we can find one that captures a lot of points. More precisely, by pigeonhole we can find $L \leq dm^2$ such that

$$|X_L| \geq \frac{m^d}{dm^2}.$$

This set is a subset of the sphere, so it doesn't have any 3-APs (it doesn't even have any points on a line).

Now we turn it into a subset of $\mathbb{Z}$ by considering a base-$2m$ expansion — we define

$$\varphi(x_1, \dots, x_d) = \sum_{i=1}^{d} x_i (2m)^{i-1}.$$

This map is injective on X (because we're only looking at digits up to m, so we don't have collisions). Even more strongly, three points $x, y, z \in \{0, \dots, m\}^d$ satisfy $x + z = 2y$ if and only if their projections satisfy the same equation — i.e., $\varphi(x) + \varphi(z) = 2\varphi(y)$. (Of course one direction is obvious; for the other, we've taken our base $2m$ to be large enough that it's not possible to 'accidentally' satisfy this equality due to carries or wrap-around.) In other words, if this equality holds in $\mathbb{Z}$, it has to be true digit-by-digit.

Then since X_L is a subset of a sphere, it is 3-AP-free; so then $\varphi(X_L)$ is a subset of $[(2m)^d]$ that is also 3-AP-free, and has size at least m^d/dm^2.

Now we've produced a subset of integers which is 3-AP-free and reasonably large; it just remains to choose the parameters. Given N, we set $m = \frac{1}{2}e^{\sqrt{\log N}}$ and $d = \sqrt{\log N}$ (or rather, the floors of these expressions) — this is asymptotically optimal — and then we get a size of $Ne^{-C\sqrt{\log N}}$. □

This shows there exist large subsets — of almost linear size — that are 3-AP-free.

> **Remark 8.4.** What did Salem and Spencer do? Here, we considered a sphere as defined by the Euclidean norm. But you can consider other convex sets. They considered a 'combinatorial sphere' where you take the same number of coordinates of each type.

This idea of looking at a high-dimensional set and projecting to $\mathbb{Z}$ in a way that preserves properties is one we'll see later, when discussing set addition (the name *Freiman homomorphism* will come up).

§8.3 Lower bounds for graph theory

So this is a set that is 3-AP-free and quite large. Last time, we proved upper bounds on 3-AP-free sets using certain graph theory results; while reversing the arguments, we also get lower bounds for those graph theory results.

In particular, we can get a lower bound for the diamond-free lemma. Recall that the diamond-free lemma says that if we can lay down triangles and we must avoid diamonds (a C_4 with one diagonal), we can't lay down too many.

> **Lemma 8.5** (Diamond-free lemma)
> In a n-vertex graph where every edge lies on exactly one triangle, there are at most $o(n^2)$ edges.

The lower bound tells us that you can get almost-quadratic numbers of triangles:

> **Corollary 8.6**
> For every n, there exists a n-vertex graph with at least $n^2 e^{-C\sqrt{\log n}}$ edges where every edge lies in exactly one triangle.

We can use the same proof as the one we used to deduce Roth's theorem from the diamond-free lemma:

Proof. Starting with a set $A \subseteq [N]$ that is 3-AP-free, last class we constructed a graph where every edge lies in exactly one triangle, and there are asymptotically $N|A|$ edges. □

This is the best lower bound we know for the diamond-free lemma. This sounds like a similar story to Roth, but in a way it's not — whereas Roth's theorem has much better upper bounds (even since Roth's work), we don't know any better bounds for this lemma than what's given by the regularity approach. So there's a giant gap between this lower bound and the upper bound we saw; it'd be very exciting to close it in either direction.

This is likewise true for many other constructions — the best-known construction for the corners theorem also comes from the same idea of Behrend's construction.

§8.4 Graph counting lemmas

One of the key ideas we saw was the triangle counting lemma, which, roughly speaking, says that if we take three parts that are pairwise regular and whose edge densities are not too small, then we can count triangles as if this were a random graph — the number of triangles between these three parts is similar to that of a random graph with the same edge densities. (We then used this in the regularity proof of the triangle removal lemma.)

But there's nothing special about triangles, and you can do this with other graphs too. We'll skip a lot of details, since they're not that interesting; but we'll explain the key ideas of how the proof works. (We'll state the theorems precisely but leave some of the details out.)

First we'll see an extension from triangles to K_4's.

Regularity can be a bit of a learning curve; part of the reason is that in combinatorics you might be used to paying attention to every word as if it matters (e.g. in Hall's theorem every word does matter). But here this isn't true — it doesn't matter if we write down ε or 2ε, for instance.

Lemma 8.7 (K_4 counting lemma)

Let $0 < \varepsilon < 1$, and suppose we have four sets X_1, X_2, X_3, and X_4 such that (X_i, X_j) is ε-regular with density $d_{ij} = d(X_i, X_j) \geq 3\sqrt{\varepsilon}$. Then the number of K_4's in $X_1 \times X_2 \times X_3 \times X_4$ is at least

$$(1-3\varepsilon)(d_{12}-3\varepsilon)(d_{13}-\varepsilon)(d_{14}-\varepsilon)(d_{23}-\varepsilon)(d_{24}-\varepsilon)(d_{34}-\varepsilon)\,|X_1|\,|X_2|\,|X_3|\,|X_4|\,.$$

We shouldn't worry about the specifics of these small constants — the hypothesis says that we have four sets which are pairwise regular and whose densities aren't too small. The conclusion says that the number of K_4's with one vertex in each set is reasonably large — at least approximately what you'd expect from multiplying the edge densities together, with a bit of room to account for imperfections.

How did the proof of the triangle counting lemma go? There we tried to count triangles by embedding vertices one at a time. We said that a typical vertex in the first set had the property it has lots of neighbors in the other two; this is a property of regularity (otherwise you'd find a witness). Then since these sets are large, there's lots of edges in between; this gives a way to embed triangles.

We can essentially do the same thing for K_4's:

Proof. Here we have four sets, and want to embed K_4. We first pick a typical vertex in X_1 — other than some small number of exceptions, this vertex will have lots of neighbors in each of the three sets. Then we repeat — we also find a typical vertex in the neighbor-set in X_2, which will have lots of neighbors in the neighbor-sets in X_3 and X_4. And then we have lots of edges between those two neighbor-sets. □

We should be convinced that this process generalizes to arbitrary graphs. There's some interesting things that lie in the details, but first we'll highlight some ways to think about this problem that might be useful.

Suppose that we're trying to count copies of C_4 with one diagonal. The counting lemma would tell us that we can embed the set in a given configuration. Here we don't care what happens between the two parts not connected by an edge.

But in some situations, $X_1, \dots, X_4$ might not actually be disjoint. This could be the case when you apply regularity and need to think about what happens inside a part. Nevertheless, it's fine to think about the multipartite sets, because we can always convert to multipartite.

To see why, suppose we talk about counting triangles in overlapping sets (we have three overlapping sets X_1, X_2, and X_3, and we want to embed triangles here). We can then create an auxiliary graph with disjoint vertex sets and the same edge relations — if we had a vertex in the intersection of X_1 and X_2, we make two copies of it and put one into our copy of X_1, and one in our copy of X_2.

So it's always fine to think about the multipartite version of the problem when discussing counting lemmas.

In other words, we never have to require that $X_1, \dots, X_4$ are disjoint — but this isn't a huge inconvenience, because you can go through this process where you create an auxiliary graph that is multipartite (by creating a graph with disjoint sets, and only drawing edges between parts). Then triangles in the original grpah correspond to triangles in the multipartite set.

In the multipartite setting, some things are also easier to argue. For example, sometimes we want to consider counting *induced* subgraphs. For example, a *seagull* is a graph consisting of two edges sharing a vertex; if we count (induced) seagulls, we need the third edge to not be present.

In the multipartite case, we can reduce this to a problem we already know — we can take the complement of all the edges between X_1 and X_3, and count triangles in this graph (and then everything is the same). This is why pulling the vertices apart is helpful (since if the three sets were the same, then this operation is less natural to think about).

Let's also give a version of the graph counting lemma which is more general and has an important feature.

Theorem 8.8 (Graph counting lemma)

Let H be a graph with maximum degree $\Delta \geq 1$, and with $c(H)$ connected components. Let $\varepsilon > 0$, let G be a graph, and let $X_i \subseteq V(G)$. Suppose that for every $ij \in E(H)$, (X_i, X_j) is ε-regular with edge density $d_{ij} = d(X_i, X_j) \geq (\Delta + 1)\varepsilon^{1/\Delta}$. Then the number of graph homomorphisms from H to G where each vertex $i \in V(H)$ is mapped to its corresponding set $X_i \subseteq V(G)$ is at least

$$(1 - \Delta\varepsilon)^{c(H)} \prod_{ij \in E(H)} (d_{ij} - \Delta\varepsilon^{1/\Delta}) \cdot \prod_{i \in V(H)} |X_i| \,.$$

Furthermore, if $|X_i| \geq v(H)/\varepsilon$ for all i, then there exists such a homomorphism $H \to G$ that is injective (i.e., a subgraph of G isomorphic to H).

Let's digest this statement. In a rough form, it's an extension of everything we discussed so far (we shouldn't worry about the details of the constants). We have a bunch of sets, and pairs of them are regular and not too sparse; and then the number of embeddings is what you'd expect from a random graph.

But there's an important difference to what we've seen so far. The point is that the dependencies here only rely on the maximum degree of H; this will be important for some applications (where the dependencies in the counting lemma need to depend only on the maximum degree, and not the size). If we'd run a naive version of this argument, we might at each step cost something proportional to the size of H; but it turns out we can do things more efficiently.

Why should this even be plausible? We have some graph; imagine it doesn't have very large maximum degree (even if it has a lot of vertices), and we're trying to embed H into some regular partition. The process is similar to what we've said before — at each step we think about how to embed each vertex of

H. And at every step of the game, we need to worry about, am I making a mistake (taking a non-typical vertex, which will cause us to get stuck later)? How many things do I need to check to make sure that I'm not stuck? When I've embedded 1, 2, 3, 4 and I'm up to embedding 5, I need to look ahead — it could be that maybe the edges of 5 are going to some additional vertices. I need to make sure that the requirements for my next vertex are met, and I haven't reduced the eligible sets for forward vertices by too much — and that's done by checking the Δ neighbors remaining. (Each step I need to check some requirements, and there are Δ vertices we need to check for.)

§8.5 Some applications

We've developed the regularity lemma (which gives a partition); and its counterpart is counting. We used the triangle counting lemma to deduce the triangle removal lemma, and similarly, using the graph counting lemma we can prove a graph removal lemma.

Lemma 8.9

For every H and ε, there exists δ (depending on H and ε) such that any n-vertex graph G with fewer than $\delta n^{v(H)}$ copies of H can be made H-free by removing fewer than εn^2 edges.

This is similar to the triangle removal; it says that if you have few copies of H, you can make the graph H-free by removing few edges. The proof is basically verbatim to the proof of the triangle removal lemma, so we won't go through it.

This also has an application to the Erdős–Stone–Simonovits theorem (which we saw in the first chapter):

Theorem 8.10 (Erdős–Stone–Simonovits)

We have

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

Last time, we saw a proof using Turán's theorem, supersaturation, and KST. Now we'll see another proof using the regularity lemma. This will be another application of the regularity recipe, which has three steps:

(1) *Partition* the graph using the regularity lemma.

(2) *Clean* the graph to deal with parts that are not so nice.

(3) Apply a *counting* lemma to get the desired result.

Proof. Fix $\varepsilon > 0$, and suppose that G is a n-vertex graph with

$$e(G) \geq \left(1 - \frac{1}{\chi(H) - 1} + \varepsilon\right) \frac{n^2}{2}$$

edges. We want to show that if n is large enough, then the above properties always guarantee that G contains H as a subgraph.

Let's apply the regularity recipe. First we apply the regularity lemma to get a partition of G into m parts $V = V_1 \cup \cdots \cup V_m$ that is η-regular, for some small η we will decide later on (where m is a constant that only depends on η).

Now we do cleaning — we remove an edges $(x, y) \in V_i$ if any of the following holds:

(a) (V_i, V_j) is not η-regular;

(b) $d(V_i, V_j) < \frac{\varepsilon}{8}$;

(c) $\min(|V_i|, |V_j|) < \frac{\varepsilon n}{8m}$.

We want to count how many edges we've removed. The number of edges removed in (a) is at most $\eta n^2 \le \varepsilon n^2/8$ (if we choose η small enough). The number of edges removed in (b) is at most $\varepsilon n^2/8$ (because between each part the density is at most $\varepsilon/8$). Likewise, the number of edges removed in (c) is at most $m \cdot \frac{\varepsilon n^2}{8m} \le \varepsilon n^2/8$.

So we've removed a fairly small number of edges — we've removed no more than $3\varepsilon n^2/8$ edges.

Let G' be the resulting graph. Then G' still has a lot of edges — it still has at least

$$\left(1 - \frac{1}{\chi(H)-1} + \frac{\varepsilon}{4}\right)\frac{n^2}{2}$$

edges. This number is bigger than the number coming from Turán's theorem, so by Turán's theorem G' contains a copy of $K_{\chi(H)}$.

In your head, imagine $\chi(H)$ being 3, so we get a triangle. The graph we're trying to embed is not a triangle, but it's 3-colorable; so we can partition it into 3 vertex sets based on the color classes. Now G' contains a copy of $K_{\chi(H)}$, which will lie between some three vertex sets. But G' was a cleaned-up graph (where we got rid of all the irregular parts and edges between sparse parts). So now we can map the vertices of H in the three color classes into these three corresponding parts, and apply the counting lemma to embed H into G.

At this point, we need to say a bunch of words about choosing the parameters to be small enough; but if you do so, then the argument works.

And we've found a copy of H in G', and therefore G; so we're done. □

What happened here? This proof still assumes Turán's theorem. Here, first you do regularity. Then suppose you want to embed a blowup of a triangle (or a subset thereof). What you do is you first find a copy of a triangle using Turán's theorem. Then using regularity and the counting lemma, you say that if you can find a copy of a triangle, you can also find any copy of a 3-partite structure (you have lots of edges between these parts, or else you couldn't find a triangle; so you can use the counting lemma). So this gives a way to boost a triangle into a larger structure.

It's good to ponder on this proof and think about what happened. We needed regularity for the last step — just having a triangle wouldn't otherwise let you power it up into bigger structures.

§8.6 Application to property testing

We'll now discuss a computer science interpretation of the regularity approach.

This concerns a class of problems in computer science known as *property testing*. The idea is that you have a very large dataset, and you only have access to a small part (e.g. a random sample). Can you conclude something about the original data?

We'll look at a specific version for graphs.

> **Question 8.11.** Suppose we're given a large (and fairly dense) graph G, and we get to sample some vertices and look at the edges between them. We want to answer, is this graph triangle-free?

In a way, this is a very hard question — maybe there's one triangle hidden somewhere in the graph, and you'll never see those vertices. So this question is impossible to answer as is. Instead, we want to distinguish two possibilities — triangle-free vs. far from triangle-free.

Question 8.12. Suppose that we instead want to distinguish two possibilities — either the graph is triangle-free, or it's ε-far from triangle free (i.e., we need to remove more than εn^2 edges to make the graph triangle-free).

So we have a graph that's either triangle-free or robustly not triangle-free; and we have to distinguish these two possibilities. (One version of the problem is that you're given the graph and you know *a priori* one of the two is the case; another version is that you have to give the right answer in one of the two cases but are not liable for your answer otherwise.)

There's a very simple algorithm:

Algorithm 8.13 — Sample k vertices uniformly at random. Check if there exists a triangle among these k vertices. If no, then output triangle-free; if yes, output ε-far-from-triangle-free.

This is a very naive algorithm; the following theorem tells us there's some probabilistic guarantees on its behavior.

Theorem 8.14

(1) If G is triangle-free, then the algorithm always outputs triangle-free.

(2) If G is ε-far from being triangle-free, the algorithm outputs ε-far from triangle fre ewith probability at least 0.99.

(3) For other G, we make no promises.

First, (1) is obvious — if the graph is triangle-free, you never see a triangle, so you always output no.

In (2), we want to output that it's ε-far from triangle-free; but we might have gotten unlucky and not seen a triangle. So we really have a probabilistic guarantee (the constant 0.99 can be replaced with anything).

Remark 8.15. This is sometimes called a *one-sided tester* because it makes a deterministic guarantee in one direction, and a probabilistic one in the other. You can boost 0.99 to an arbitrarily high probability by repeating the test many times.

Proof. The main point is that this claim is basically equivalent to the triangle removal lemma. Let's think about why — suppose that G is ε-far from triangle-free. This means we need to remove a lot of edges to make it triangle-free. The triangle removal lemma says that if there's few triangles, then we can make it triangle-free by removing few edges; so contrapositively, if it's ε-far from being triangle-free, then it must have at least δn^3 triangles. So if we sample triples of vertices, we should with high probability see one of these triangles; and that's the entirety of this claim. □

This is the property of containing a triangle; but what about other graph properties (subsets of graphs that are invariant under isomorphisms, i.e., don't depend on the labelling of the vertices)?

Definition 8.16. A graph property is *hereditary* if it's closed under vertex deletion.

For example, planarity is hereditary (since if a graph is planar and we delete a vertex, it's still planar). Being triangle-free is also hereditary.

Question 8.17. What about property testing for an arbitrary hereditary property?

It turns out that the same thing is true:

Theorem 8.18

Let $\mathcal{P}$ be any hereditary property, and fix $\varepsilon > 0$. Then there exists k only depending on $\mathcal{P}$ and ε such that the following holds: we're asked whether $G \in \mathcal{P}$, or whether it's ε-far from $\mathcal{P}$ (i.e., we can add or remove at most εn^2 edges to reach $\mathcal{P}$). Then the same algorithm holds, with the same guarantees.

We haven't yet developed the tools to prove this theorem. The reason is here we get to not only remove edges, but also add edges; that's different from what we proved. And this is an important difference:

Example 8.19

A graph having no *induced* copies of H is a hereditary property.

This requires an *induced* graph removal lemma. We'll do this next time, but it's important to realize why we haven't yet proved it. We proved a *graph* removal lemma, but not an *induced* graph removal lemma — why? You start with a graph that has a few copies of H, and you want to get rid of all induced copies. But suppose we do the cleaning step where we get rid of all these edges. In an extreme case H may be the empty graph; then this certainly doesn't help us get rid of copies of H. So getting rid of edges might hurt you — you might actually end up with *more* copies of H. Next time we'll develop a stronger notion of regularity that'll allow us to prove the induced graph removal lemma.

§9 October 4, 2023

We spent three lectures discussing the graph regularity method — we first proved Szemerédi's graph regularity lemma about partitioning an arbitrary graph into sets that are mostly pairwise regular. Then we proved the triangle counting lemma, and used it together with the regularity recipe to prove the triangle removal lemma. Last class we looked at Behrend's construction for large 3-AP-free sets, and proved the graph removal lemma; and we discussed property testing towards the end of the class.

When we were discussing property testing, we saw that the problem of trying to distinguish a triangle-free graph from a graph that is very far from triangle-free has a very simple algorithm — sample and then test — and the proof of correctness is basically equivalent to the triangle removal lemma. But for other properties, namely containing an induced graph, we have not proved such a lemma. We'll do that today — we'll prove an induced version of the graph removal lemma.

§9.1 Induced graph removal lemma

Theorem 9.1 (Induced graph removal lemma)

For every H and $\varepsilon > 0$, there exists $\delta > 0$ such that if a n-vertex graph has fewer than $\delta n^{v(H)}$ induced copies of H, then it can be made induced-H-free by adding or deleting (or both) at most εn^2 edges.

Recall that an *induced* copy means that all edge and non-edge relations need to be there (as opposed to a subgraph, where we permit additional edges to be present).

What's the difficulty here? In the regularity recipe, we had three steps:

(1) Partition (using the regularity lemma).

(2) Perform a cleaning step. In the applications we saw, we removed irregular pairs, low-density pairs, and pairs involving small vertex parts.

(3) Use a counting lemma.

What's easy and what's hard? We discussed last time that counting for induced subgraphs is the same as counting as we've seen — we can first go to a multipartite graph, and then if we want to count embeddings of an induced two-edge path, we can complement the third part; and then it becomes the same thing as counting triangles. So there's no additional ideas involved in counting, at least for now.

Small parts were never really an issue — if we use the equitable version of regularity then there are none (all parts have roughly the same size).

What about low-density pairs? Before this, note that it's important that here we allow both adding *and* deleting edges. For example, suppose we're trying to forbid induced copies of the empty 3-vertex graph, in a graph G which is a complete graph with a triangle removed. This has exactly one copy of our empty graph H, but if we only delete edges we can't get rid of it. So we can't only delete; we must also allow adding. (For induced graphs there's a symmetry between them.)

Similarly here, we have a symmetry between low-density pairs and high-density pairs. If the density is less than ε, we should instead delete the edges, and if the density is greater than $1-\varepsilon$ we should add the edges.

But the kicker is irregular pairs. We can't get rid of irregular pairs, and what should we do with them? Previously, we just removed the edges. But that no longer makes sense in the context of induced graph removal — if you remove edges, you still have to worry about what happens between those two vertex sets (they don't just go away, and you might still be able to embed). So this step is the difficulty.

On one hand, we cannot get rid of irregular pairs — they have to be there. On the other hand, we don't quite know what to do with them in the context of proving induced graph removal. So that's the challenge.

Let's see how we can overcome this challenge. We'll do this by coming up with a stronger notion of regularity — we'll enhance our regularity lemma in order to prove what we want.

§9.2 Strong regularity lemma

Here is a version of the *strong regularity lemma*. There are lots of notions that go by similar names (we'll see two things with this name, and use one to prove the other).

Recall the notion of energy — we defined the energy $q(\mathcal{P})$ in our proof of the regularity lemma. It's essentially defined as an expectation of squared density between parts in our partition; this was an important notion when we proved the regularity lemma.

The strong regularity lemma says the following. In the Szemerédi regularity lemma there was a single constant as an input; here we have a sequence. You can think of ε_k as $\varepsilon/2^k$ if you like (some concrete decreasing function).

Lemma 9.2 (Strong regularity lemma)

Given any $\varepsilon_0 \geq \varepsilon_1 \geq \cdots$, there exists M (only depending on this sequence) such that every graph has a *pair* of partitions $\mathcal{P}$ and $\mathcal{Q}$ such that the following is true:

(a) $\mathcal{Q}$ refines $\mathcal{P}$.

(b) $\mathcal{P}$ is ε_0-regular.

(c) $\mathcal{Q}$ is $\varepsilon_{|\mathcal{P}|}$-regular.

(d) $q(\mathcal{Q}) \leq q(\mathcal{P}) + \varepsilon_0$.

(e) $|\mathcal{Q}| \leq M$.

The point (c) is saying that $\mathcal{Q}$ is *extremely* regular — think of ε_i as a sequence that's rapidly decreasing. We do the regularity partition and find some $\mathcal{P}$, and $\mathcal{P}$ will have a bounded number of parts, but lots of them. And then if it has a lot of parts, $\mathcal{Q}$ will have to obey an even stricter regularity condition.

For (d), recall that $\mathcal{Q}$ refining $\mathcal{P}$ means that $q(\mathcal{Q}) \geq q(\mathcal{P})$ (refinement never decreases energy). But (d) says that the energy of $\mathcal{Q}$ doesn't differ too much from that of $\mathcal{P}$. In a sense, this says that $\mathcal{Q}$ and $\mathcal{P}$ are not that different (they're close in some sort of L^2-distance).

This is a notion of a strong regularity lemma — it produces not just one partition, but a pair of partitions.

The proof is actually not that hard; in fact the proof might be a way to digest what's happening in the theorem. But it uses a nice idea we kind of already saw — the basic idea is that we're going to apply Szemerédi's regularity lemma repeatedly. We start with some ε-regular partition $\mathcal{P}_0$, and then refine it further to get another partition $\mathcal{P}_1$, and then $\mathcal{P}_2$, and so on — this gives a sequence of refinements $\mathcal{P}_0 \rightsquigarrow \mathcal{P}_2 \rightsquigarrow \mathcal{P}_3 \rightsquigarrow \cdots$. And because the energy is between 0 and 1, it must increase by less than ε_0 at some step; if this happens on the step $\mathcal{P}_2 \rightsquigarrow \mathcal{P}_3$, then we choose these as $\mathcal{P}$ and $\mathcal{Q}$.

We'll use the following version of Szemerédi's regularity lemma (which has the same proof):

Theorem 9.3

For all $\varepsilon > 0$ and k, there exists $M_0 = M_0(k, \varepsilon)$ such that for all partitions $\mathcal{P}$ of $V(G)$ with $|\mathcal{P}| \leq k$, there exists a refinement $\mathcal{P}'$ of $\mathcal{P}$ such that every part of $\mathcal{P}$ is refined into at most M parts, and such that $\mathcal{P}'$ is ε-regular.

The version of regularity that we proved is the one where we start with the trivial $\mathcal{P} = \{V(G)\}$; then a refinement is just a partition, and we found one that is ε-regular and has a bounded number of parts. But our proof involved starting with the trivial partition and repeatedly cutting it up; if we start with a different partition $\mathcal{P}$, the same proof carries through to get this result. So we can use the same proof, just starting with $\mathcal{P}$ instead of the trivial partition.

Now we'll use this to carry out the above strategy.

Proof. We start with the trivial partition $\mathcal{P}_0 = \{V(G)\}$, and for each i, we define $\mathcal{P}_{i+1}$ to be obtained by applying this version of Szemerédi's regularity lemma, so that $\mathcal{P}_{i+1}$ refines $\mathcal{P}_i$ and is $\varepsilon_{|\mathcal{P}_i|}$-regular. (At the ith step we know the number of parts in $|\mathcal{P}_i|$, so we can pick this constant $\varepsilon_{|\mathcal{P}_i|}$; and then we can use this to refine $\mathcal{P}_i$ into a partition with specified regularity.)

Since $0 \leq q(\mathcal{P}_i) \leq 1$ and energy only increases, there must exist some $i \leq \varepsilon_0^{-1}$ such that $q(\mathcal{P}_{i+1}) \leq q(\mathcal{P}_i) + \varepsilon_0$ (we can't increase by ε_0 too many times). And then we're done, since this is exactly what we're looking for — we take $\mathcal{P}_i = \mathcal{P}$ and $\mathcal{P}_{i+1} = \mathcal{Q}$, and all the conditions are satisfied. (The number of parts in $\mathcal{Q}$ is bounded because it's bounded at each step, and we have a bounded number of steps.) □

Remark 9.4. We only required $\mathcal{P}$ to be ε_0-regular; in reality it might be a lot more regular. But we don't actually need that — when you apply this you set ε_0 to be what you're looking for, and go from there.

Remark 9.5. What kind of quantitative bounds does this give? Each application of Szemerédi's regularity lemma produces tower-type bounds. If we suppose that $\varepsilon_i = \varepsilon/\mathsf{poly}(i)$, then the first time, the number of parts ends up being $\mathsf{tow}(\mathsf{poly}(\varepsilon^{-1}))$. Then this is the number of parts, so it goes into the denominator — so this becomes roughly the size of ε^{-1}. And then we have to do regularity again; so we get

$$\mathsf{tow}(\mathsf{tow}(\mathsf{tow}(\cdots(\mathsf{poly}(\varepsilon^{-1}))))),$$

where the tower is repeated ε_0^{-1} times. So we're going one step up in the Ackerman hierarchy of functions; in the combinatorics community this is called *wowzer* growth. (You can tell whether someone is an extremal combinatorialist by whether they recognize this word — *wowzer* is as in 'wow, that's a really big number'.)

But we won't worry about quantitative bounds for this lecture — all we're concerned with is existence.

Remark 9.6. We can obtain equitability here as well, by an appropriate modification of this proof where you maintain equitability at each step (we'll use this later to not worry about small parts).

§9.3 Interpreting the energy condition

The next thing we'll talk about is — how do we understand the condition of our two partitions having close energies to each other?

Lemma 9.7

Suppose that $\mathcal{P}$ and $\mathcal{Q}$ are both vertex partitions of G, with $\mathcal{Q}$ refining $\mathcal{P}$. For each vertex $x \in V(G)$, write V_x to be the part of $\mathcal{P}$ that x lies in, and W_x to be the part of $\mathcal{Q}$ that x lies in.

If $q(\mathcal{Q}) \le q(\mathcal{P}) + \varepsilon^3$, then

$$|d(V_x, V_y) - d(W_x, W_y)| \le \varepsilon$$

for all but at most εn^2 pairs of vertices $(x, y) \in V(G)^2$.

So we have a $\mathcal{P}$-partition, and its refinement is the $\mathcal{Q}$-partition. And we pick a random vertex x and ask, which cell in $\mathcal{P}$ does it lie in, and which cell in $\mathcal{Q}$ does it lie in? (These are V_x and W_x, respectively.)

This states that if we pick two vertices uniformly at random and look at their corresponding vertex sets in the coarse partition and also the fine partition, and look at how their densities differ — we want to say that the edge densities between parts in $\mathcal{Q}$ should roughly be the same as those in $\mathcal{P}$, and having energies close together more or less guarantees that (up to a small number of exceptions).

Remark 9.8. Here we don't require any regularity; we only require that the energies are close together.

Proof. The proof highlights the idea that energy is some L^2-quantity; so somehow being close in energy is some way of capturing being close in Euclidean distance.

Let x and y be uniformly chosen in $V(G)$. As in the proof we saw last Monday for the regularity lemma, we define $Z_{\mathcal{P}} = d(V_x, V_y)$, and likewise $Z_{\mathcal{Q}} = d(W_x, W_y)$. We saw that

$$q(\mathcal{P}) = \mathbb{E}[Z_{\mathcal{P}}^2] \text{ and } q(\mathcal{Q}) = \mathbb{E}[Z_{\mathcal{Q}}]^2.$$

(This is essentially the definition of energy.)

Claim 9.9 — We have

$$q(\mathcal{Q}) - q(\mathcal{P}) = \mathbb{E}[Z_{\mathcal{Q}}^2] - \mathbb{E}[Z_{\mathcal{P}}^2] = \mathbb{E}[(Z_{\mathcal{Q}} - Z_{\mathcal{P}})^2].$$

(The first equality is the definition; the second is the claim.)

This should be familiar in other contexts — it's the Pythagorean theorem. Namely, imagine a right triangle with $Z_{\mathcal{P}}$ and $Z_{\mathcal{Q}} - Z_{\mathcal{P}}$ as legs, and $Z_{\mathcal{Q}}$ as the hypotenuse.

Proof. This identity is equivalent to stating that $\mathbb{E}[Z_p(Z_{\mathcal{Q}} - Z_{\mathcal{P}})] = 0$ (i.e., that we actually have a right angle). First, $Z_{\mathcal{P}}$ is what happens if we pick two vertices x and y, and look at their coarse parts. If x and y vary inside their own parts V_x and V_y, then Z_p is constant (by definition).

But now we can look at $\mathcal{Q}$. As x and y vary over their own cells in $\mathcal{P}$, what happens to $Z_{\mathcal{Q}}$? It gets averaged out to precisely $Z_{\mathcal{P}}$. So when this happens, $Z_{\mathcal{Q}} - Z_{\mathcal{P}}$ averages to 0 (in each cell). And that's the proof. □

So having a bound on the energy difference implies that $Z_{\mathcal{Q}} - Z_{\mathcal{P}}$ is small in L^2-norm — i.e., we now have

$$\mathbb{E}[(Z_{\mathcal{Q}} - Z_{\mathcal{P}})^2] \le \varepsilon^3.$$

In particular, by Markov's inequality

$$\mathbb{P}(|Z_{\mathcal{Q}} - Z_{\mathcal{P}}| > \varepsilon) \le \varepsilon.$$

But this is precisely what we needed to prove. □

So that's an interpretation of the energy condition — having two partitions, *one refining the other* (this is important) and their energies being close to each other tells us that their densities are also very close to each other. So in that sense, these two partitions are quite close to each other.

§9.4 Another strong regularity lemma

Let's use this to prove another notion of a strong regularity lemma.

Lemma 9.10

Suppose that $\varepsilon_0 \ge \varepsilon_1 \ge \cdots$ is a sequence of decreasing constants. Then there exists $\delta > 0$ such that every graph has an equitable partition $V_1 \cup \cdots V_k$ into k parts, as well as a distinguished subset $W_i \subseteq V_i$ for each i, such that:

(a) $|W_i| \ge \delta n$.

(b) (W_i, W_j) is ε_k-regular for *all* $1 \le i \le j \le k$.

(c) We have $|d(V_i, V_j) - d(W_i, W_j)| \le \varepsilon_0$ for all but at most $\varepsilon_0 k^2$ pairs $(i, j) \in [k]^2$.

So we have a partition, and inside each part we have a distinguished W_i.

Note that (a) bounds the number of parts automatically.

The important thing in (b) is that it's true for *all* the pairs, with no exceptions — between the parts as well as inside each part, all are ε_k-regular.

Let's discuss the conclusion of this lemma. At the beginning of today's class, we said one difficulty in proving the induced graph removal lemma is that there could be irregular pairs; that's unavoidable. This lemma says that we're not going to try to guarantee that all the pairs themselves are regular; but if I only get to pick a large representative subset from each V_i, *then* I can guarantee that all the representatives W_i are

pairwise regular. So we don't have the whole graph, but a large representative from each piece. This will be important for us because we really don't want irregular pairs, and this will be a good way to get something with no irregular pairs.

We'll now sketch a proof of the strong regularity lemma as above (but we won't fully prove it). We'll only prove a weaker statement, where in (b), we only consider $i < j$; we won't guarantee that each W_i is regular with itself. This is actually on our problem set — it's not precisely what's written here, but if we figure out what's happening in the problem set, we should be able to handle each part with itself.

Proof sketch. We'll apply the strong regularity lemma. First note that we can always decrease ε_i if needed (if we make them smaller, we're just proving something stronger); sometimes we'll want them to be small.

First, we apply the earlier version of strong regularity, which gives us a pair of partitions where the energy between them is quite small. We'll assume everything is equitable (for convenience; it's important for this proof). So it gives us:

- An equitable ε_0-regular partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of G;
- An equitable ε_k-regular partition $\mathcal{Q}$ refining $\mathcal{P}$, such that $q(\mathcal{Q}) \le q(\mathcal{P}) + \varepsilon_0^3/8$.
- We have $|\mathcal{Q}| \le M$.

In each V_i, choose a uniform random part W_i of $\mathcal{Q}$. (We've started with $\mathcal{P}$, which gives some V_i; then inside each piece, $\mathcal{Q}$ partitions it further, and we pick one of these pieces at random.) Equitability guarantees us that each set is not too small ($|W_i| \ge \delta n$, for $\delta \approx 1/M$).

Since $\mathcal{Q}$ is ε_k-regular, we know that all but an ε_k-fraction of pairs of parts of $\mathcal{Q}$ are ε_k-regular. The point here is that we get to choose ε_k as an arbitrarily small function of k. And if we have V_1 and V_2 and we choose W_1 and W_2 at random, $\mathcal{Q}$ is such that it's unlikely for (W_1, W_2) to be irregular — the *expected* number of pairs (W_i, W_j) that are *not* ε_k-regular is at most $\varepsilon_k k^2$ (the k^2 is for pairs of (i, j), and for each (i, j) we have at most this probability of hitting a bad pair).

So with positive probability, e.g. probability at least 90%, (W_i, W_j) is ε_k-regular for all $i < j$.

There's also a final thing we need, which is that the densities between V_i and V_j are similar to that of W_i and W_j. This follows from the fact we just proved (using the similarity in energies). □

Remark 9.11. Why doesn't this approach give us $i = j$ — why doesn't this show that W_i is ε_k-regular? Nothing so far precludes the possibility that *all* parts of $\mathcal{Q}$ are irregular with themselves. And there's no cleverness in this part of the argument that can overcome this issue if that happens. But if you look at two parts, that's a different story.

§9.5 Induced graph removal lemma

Now let's assume this version of the strong regularity lemma; and we'll prove the induced removal lemma. This is similar to what we did before, but using this stronger notion.

Proof. We first apply the strong regularity lemma to get a partition $V_1 \cup \cdots \cup V_k$, as well as a representative $W_i \subseteq V_i$.

We now do a cleaning step. First, what are the guarantees of the partition? We have that (W_i, W_j) is ε'-regular for all $i \le j$ (where ε' is a small constant we choose), and that $|d(V_i, V_j) - d(W_i, W_j)| \le \varepsilon/8$ most of the time — i.e., for all but at most $\varepsilon k^2/8$ pairs. (We shouldn't worry about the random constants.) And finally, we have $|W_i| \ge \delta_0 n$ for some constant δ_0.

So this is the output of the strong regularity lemma. We now perform the cleaning step. It has two parts.

First, for all $i \le j$ (including $i < j$):

(a) If $d(W_i, W_j) \le \varepsilon/8$, then we remove all edges in (V_i, V_j).

(b) If $d(W_i, W_j) \ge 1 - \varepsilon/8$, then we add in all edges in (V_i, V_j).

Note that in the cleaning step, we are not simply fixing the edges between W_i and W_j; we're looking at what happens between W_i and W_j, and using that to tell us what to do with *all* the edges between V_i and V_j. Compared to earlier, there's no clause about small parts because we're using equitable partitions. But more importantly, there's no clause about irregular pairs, because there are none; so we don't have to face the dilemma where for irregular pairs we don't know what to do.

We can now bound the number of edges changed in this process — for all but a small number of exceptions, if the W_i's have small density then so do the V_i's, so we haven't changed too many edges.

Finally, we need to do counting. First, we claim that the resulting G' is induced-H-free.

Let's suppose that it's *not* induced-H-free — we've done all this cleaning, and eventually we get G', and someone finds a copy of H. For example, let's suppose H is the seagull; then this means there's some V_1, V_2, V_3 for which we have a copy here. That means, looking at the W's, that the edge densities between the W's are not too small between V_1 and V_2 and between V_1 and V_3, and it's not too large between V_2 and V_3 (otherwise we'd have changed the V's to prevent an induced copy).

Then we can use the counting lemma between the W_i's to get us lots of copies of induced H. This would violate the hypothesis of having few induced H's to begin with. □

Remark 9.12. Here, since in G' there exists an edge in (V_1, V_2), that tells us that $d(W_1, W_2)$ was originally not too small, because if this were violated, we would have deleted all the edges between them. So the W's (which are part of the original graph) have enough edge densities to perform the counting lemma.

§9.6 Infinite graph removal lemma

We'll now present an extension, called the *infinite* graph removal lemma (also induced). The reason it's interesting is it completes the claim mentioned at the end of last class — about testing arbitrary graph properties. The point is that a hereditary property is defined by forbidding some (possibly infinite) collection of induced subgraphs. (This is worth thinking about.) For some important classes of hereditary properties, such as planarity, we know exactly the set of graphs we need to forbid. But in general, by definition we can look at the minimal elements of what's not in the set.

Lemma 9.13

For every $\mathcal{H}$ (possibly infinite) and every ε, there exists h_0 and $\delta > 0$ such that if G is a n-vertex graph with fewer than $\delta n^{v(H)}$ induced copies of H for every $H \in \mathcal{H}$ with at most h_0 vertices, then G can be made induced-$\mathcal{H}$-free by adding or removing no more than εn^2 edges.

It may take some time to process the statement — why do we have h_0 here? One way to think about it is that in the application to property testing, you only get to test a finite number of vertices; so we need a cap, and the point is it's only enough to test up to that cap. If $\mathcal{H}$ is finite, this is basically identical to the induced graph removal lemma (excluding two graphs is no more difficult than one). With infinite families, you have to do more work; it's not difficult, but it's intricate and we won't do it here.

So for every $\mathcal{H}$ and every ε, to make the whole graph $\mathcal{H}$-free it suffices to verify this condition. (It's kind of subtle what the lemma says.)

§9.7 Hypergraph regularity

The final thing in this chapter concerns hypergraph regularity.

We used the Szemerédi regularity lemma to prove the triangle removal lemma, which then implied Roth's theorem. If you want to follow the same strategy and prove Szemerédi's theorem for 4-term arithmetic progressions, you'll need a *hypergraph* version — a *tetrahedron removal lemma*.

In a 3-uniform hypergraph, the edges are triples; a tetrahedron is 4 vertices with all triples present (think of the faces of a tetrahedron). And to prove a tetrahedron removal lemma, you need a *hypergraph* regularity lemma.

Graph regularity was proven in the 1970s, but hypergrpah regularity is much harder, and it was only completed in the 2000s. We won't be able to discuss even the definition of hypergraph regularity, but we'll try to convey some flavors.

First, let's talk about the second step — going from tetrahedron removal to 4-APs — which is easier. Recall that we had a diamond-free lemma; extended to hypergraphs this corresponds to the following statement.

Corollary 9.14

If G is a n-vertex 3-graph where every edge (triple) is contained in a unique tetrahedron, then G has $o(n^3)$ edges.

The statement is a straightforward generalization of the diamond-free lemma, though the proof is much more difficult. But starting from this statement, we can derive Szemerédi for 4-APs.

In the proof of Roth's theorem, we set up a graph such that the edges between the parts were determined by some equations, so that triangles correspond to 3-APs. We'll do the same for 4-APs, where tetrahedra correspond to 4-APs.

There are several ways to do this; here's one.

Proof. We'll have four sets W, X, Y, and Z, which consist of $\mathbb{Z}/M\mathbb{Z}$ (similar to Roth's theorem). We start with $A \subseteq [N]$, and we embed it in $\mathbb{Z}/M\mathbb{Z}$ where M is around a constant multiple of N (but also chosen to get rid of divisibilities). Then we draw the following edges:

- $wxy \in E$ if and only if $3w + 2x + y \in A$.
- $wxz \in E$ if and only if $2w + x - z \in A$.
- $wyz \in E$ if and only if $w - y - 2z \in A$.
- $xyz \in E$ if and only if $-x - 2y - 3z \in A$.

This looks slightly different from what we did for Roth's, but the idea is similar. The reason we chose these expressions is that these four expressions form a 4-AP, with common difference $-w - x - y - z$.

What are the tetrahedra in this graph? They correspond to 4-APs. But we don't have any 4-APs, so they must correspond to trivial 4-APs. Then the rest of the story is the same as with Roth's theorem — having 4-AP-free A implies the hypothesis in the corollary, that every edge is in a unique tetrahedron. □

We can't prove this corollary, but we'll sketch some ideas that show what the difficulties are — why is hypergraph regularity so hard?

You can try to come up with easier versions of regularity. Here's one reasonable notion of hypergraph regularity:

Definition 9.15 (Initial attempt at 3-graph regularity)**.** Given $V_1, V_2, V_3 \subseteq V(G)$, we say that (V_1, V_2, V_3) is ε-regular if for all $A_i \subseteq V_i$ with $|A_i| \geq \varepsilon |V_i|$ for all i, we have

$$|d(V_1, V_2, V_3) - d(A_1, A_2, A_3)| \leq \varepsilon.$$

This is the straightforward generalization for what we saw for graphs (where density is defined as the number of edges divided by the product of part sizes, as before).

This definition is perfectly fine, and if you start with this definition and try to prove a regularity lemma, the proof works (there's no new ideas needed). So this does lead to *a* regularity lemma.

But the point of a regularity lemma is that we can use it to prove things. And is this regularity lemma strong enough to prove our corollary?

In the regularity recipe, we needed not only a partition lemma, but also a counting lemma. So is there a tetrahedron *counting* lemma?

We'll demonstrate that for this version of regularity, the answer is definitively no — this notion is not strong enough to count tetrahedra. To see this, we'll see two hypergraphs that are very similar in this sense, but that have very different tetrahedra counts.

First, take two constants p and q in $(0, 1)$. We first build a random graph $G^{(2)} = G(n, p)$, where every pair is included as an edge with probability p. Then we construct a 3-uniform hypergraph $G^{(3)}$ by including each triangle in $G^{(2)}$ as an edge with probability q. (So we first look at a random graph, which has some number of triangles; we keep this triangle as a triple with probability q.) Let's call this hypergraph X.

Now here's a second, much simpler, notion of a random hypergraph — we simply include each triple with probability $p^3 q$.

We'll see two things: first, with respect to the above version of regularity, these two graphs look roughly the same. The two graphs have the same edge densities, and because of concentration, they'll have the same density even restricted to subsets.

In the first graph, the edge density is $p^3 q$, while the tetrahedron density is $p^6 q^4$. But in th esecond graph, the edge density is still $p^3 q$, and the tetrahedron density is $(p^3 q)^4$.

So this notion of regularity is too weak to do counting. We need a notion of regularity that's far stronger, that would allow us to count. We won't be able to say what the notion is exactly, but here's an impressionistic picture.

We need a stronger notion of regularity that'll give us counting, and this stronger notion involves: starting with a 3-graph $G^{(3)}$, we want to first partition $\binom{V}{2}$ (the set of pairs of vertices) into *graphs* $G_1^{(2)} \cup \cdots \cup G_\ell^{(2)}$ such that $G^{(3)}$ sits in a random-like way on top of these 2-graphs.

So we have a bunch of triples, and we lay some graphs below so that the hypergraph looks random with respect to the lower 2-graphs.

And then we partition V to regularize these graphs.

So we have two steps — we first introduce a partition of the complete graph into graphs, and then we regularize those graphs. This is a very delicate process, but it's what's needed for the counting lemma.

But this makes the regularity lemma much harder to prove. There are many different notions; some have easier counting lemmas and harder regularity lemmas, and some the other way around. In many cases we don't know if these notions are quantitatively equivalent to each other.

§10 October 11, 2023

We've spent the last couple of weeks talking about the graph regularity method. In the regularity lemma, there was an important concept of ε-regularity, which intuitively says that between two sets of vertices, the

graph looks 'random-like.' We'll now explore this notion in greater depth — in what senses does a graph behave when it is random-like?

§10.1 Pseudorandomness

Before talking about graphs, we'll talk more philosophically about what it means to be pseudo-random or random-like. There is a sense of *true* randomness, in terms of a probability distribution. But in many real-world situations, things work differently. If you open your computer and type `rand` into Python, what it outputs isn't an actual random number — there's an algorithm that generates something that *looks* random, and for all intents and purposes you can't distinguish it from something that's truly random. So it's *pseudorandom* in the sense you can't distinguish its output from random output.

The prime numbers are definitely not random — you can't ask 'what's the probability 97 is prime?' But there are many ways in which the primes are distributed in a random-like way — there are theorems that roughly quantify this (such as the Riemann hypothesis).

If you think about the digits of π, they are again deterministic (not random). But the sequence looks kind of random; and if you look at statistical properties, it does look kind of random. The mathematical term for this is that a number is called a *normal* number if its digit distribution converges to the uniform one. We don't actually know if π is normal, but it's suspected to be.

All of these are examples of objects that are not random, but are *pseudorandom* in the sense that they behave *like* a random object in some sense. We'll now explore that notion for graphs.

So we'll now look at pseudorandom graphs — graphs that are random-like (think of the Erdős–Rényi random graph as our model of a true random graph) in some sense.

This is a general concept; today we'll look at a specific type of pseudorandomness, generally known as *quasirandom graphs*.

§10.2 Quasirandom graphs

Theorem 10.1 (Chung–Graham–Wilson 1989)

Fix some $p \in [0,1]$, and consider a sequence of graphs (G_n) where G_n has n vertices and $e(G) = p\binom{n}{2} + o(n^2)$. We'll write $G = G_n$. Then the following properties are equivalent:

- DISC — $e(X,Y) = p\,|X|\,|Y| + o(n^2)$ for all $X, Y \subseteq V(G)$.
- DISC — we have $e(X) = p\binom{|X|}{+}o(n^2)$ for all $X \subseteq V(G)$.
- COUNT — for all H, the number of labelled copies of H in G is $(p^{e(H)} + o(1))n^{v(H)}$.
- C_4 — the number of labelled 4-cycles is at most $(p^4 + o(1))n^4$.
- CODEG — We have
$$\sum_{u,v} |\mathrm{codeg}(u,v) - p^2 n| = o(n^3).$$
- EIG — If $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of the adjacency matrix of G, then $\lambda_1 = pn + o(n)$ and $|\lambda_i| = o(n)$ for all $i = 2, \ldots, n$.

(It's fine to only take a subsequence of integers — n doesn't have to go through all the integers.)

The first point (called DISC for *discrepancy*) should remind us of ε-regularity — it's precisely ε-regularity as $\varepsilon \to 0$. The second point says that it's enough to look at edges *within* a set, rather than between two sets.

For the third point, in a *random* graph we'd expect $p^{e(H)}n^{v(H)}$ copies of H; the condition says that we have this count, up to a small error. The fourth condition looks much weaker — instead of being about *all* graphs, it's just about C_4.

For the fifth point, $\operatorname{codeg}(u, v)$ is the number of common neighbors; in a random graph, we'd expect it to be $p^2 n$. We're summing n^2 terms that could potentially be around n, so by default we'd expect it to be cubic; the point says that it's subcubic.

Today we'll prove this theorem; but first we'll discuss these conditions and see some examples.

Definition 10.2. We say that a sequence of graphs (G_n) is *quasirandom (at density p)* if it satisfies the above conditions.

Remark 10.3. We say that a *sequence* of graphs is quasirandom, not an individual graph. Each of these conditions is an asymptotic in n, so it doesn't make sense to say whether a single graph is quasirandom — it only makes sense to talk about a sequence. We'll colloquially say that some graph is quasirandom, but when we do there's an underlying sequence (or a quantity that goes to 0 as $n \to \infty$).

Remark 10.4. What do we mean by *labelled* graphs? This means we treat the vertices as having labels, and different ways of embedding the labels are counted differently. There are $\binom{n}{3}$ triangles in a complete graph, but $n(n-1)(n-2)$ labelled triangles.

One really surprising thing in this theorem is C_4 — the condition is just about a *single* subgraph count (namely C_4), but it implies properties about *all* subsets of the graph. This is really interesting and surprising if you haven't seen it before.

If we're given a graph (or rather, a sequence) and are asked algorithmically to check whether it's quasirandom, which condition should we use? DISC would involve checking all subsets, which would take exponential effort. But C_4, CODEG, or EIG would be much faster. So there are genuine differences; but these conditions turn out to be equivalent.

Remark 10.5. We're speaking about sequences of graphs, but we could equivalently write the statements quantitatively — we could write the statements with an error parameter ε. For example, DISC(ε) would be the statement that for all $X, Y \subseteq V(G)$, we have $|e(X,Y) - p|X||Y|| \le \varepsilon n^2$.

Then we'd need to prove that the properties with error parameters all imply each other, up to changing the parameters (in fact, you only need to change them polynomially).

§10.3 Some examples

Example 10.6

The random graph $G(n,p)$ is quasirandom at density p with probability 1.

(If this weren't true, we'd be unjustified in calling the sequence quasirandom.)

The proof is not hard; you can use the Chernoff bound. For example, for any specific X and Y, we have $e(X,Y) = p|X||Y| + o(n^2)$ with very high probability, and we can just union-bound.

But this is in fact true of lots of graphs that *aren't* random.

Example 10.7

Let $p \equiv 1 \pmod 4$. The *Paley graph* is the graph on vertex set $\mathbb{Z}/p\mathbb{Z}$, with edges $x \sim y$ if and only if $x - y$ is a quadratic residue. Then the Paley graph is quasirandom.

This is not a random graph (it's a Cayley graph, and is very structured), but it turns out to satisfy all these properties.

Other examples come from using different groups — here we used $\mathbb{Z}/p\mathbb{Z}$ but had to be judicious in our choice of generators. For some other groups, you can be quite free in what generators you choose.

Example 10.8

Consider the group $\Gamma = \mathrm{PSL}(2, p)$ (2×2 matrices with determinant 1, modulo ± 1). For any generating set S with $S = S^{-1}$, every Cayley graph $\mathrm{Cay}(\Gamma, S)$ is quasirandom.

So here it doesn't matter which generators you choose; the graph will always be quasirandom.

Example 10.9

Take $S \subseteq \mathbb{F}_p \cup \{\infty\}$. Consider the graph G whose vertex set is the affine plane $\mathbb{F}_p^2$, and where $x \sim y$ if their slope lies in S. Then G is quasirandom.

So again here we have a graph whose construction is algebraic and deterministic, but which satisfies all these properties.

§10.4 Proving equivalences

We'll soon prove the equivalences of these properties. Before this, we'll see a warm-up that will illustrate useful calculation techniques.

§10.4.1 A Warm-up

Proposition 10.10

Every n-vertex graph with at least $\frac{pn^2}{2}$ edges has at least $p^4 n^4$ labelled closed walks of length 4.

(A labelled closed walk means we count walks where we specify the starting point and look at the 4 steps, and we have to return to where we start. Note that here there's no error term; this inequality is true as written.)

The difference between a labelled walk and C_4 is that in a labelled walk you're allowed to repeat vertices, whereas in a C_4 you're not; but the difference is quite negligible. (This quantity differs from the number of labelled C_4's by $O(n^3)$, which is negligible.)

The proof will be a sequence of Cauchy–Schwarz inequalities; it's useful to see it both algebraically and visually.

Proof. To find the number of closed walks of length 4, we'll essentially do a number of reflections — Cauchy–Schwarz visually is about reflecting along some line of symmetry. So we'll first take a 4-walk, and identify two opposite vertices w and y; and try to do a reflection along the line of symmetry through them. We can then write

$$\#\text{closed length-4 walks} = \sum_{w,y} \#\{x \mid w \sim x \sim y\}^2$$

(the squaring corresponds to precisely this line of symmetry).

And now we can fold this symmetry through the use of Cauchy–Schwarz — by Cauchy–Schwarz this is at least

$$\frac{1}{n^2}\left(\sum_{w,y} \#\{x \mid w \sim x \sim y\}^2\right).$$

Visually we're folding our 4-vertex loop into a 2-edge path.

This again has a line of symmetry, so we can rewrite this in terms of our middle vertex x — it's

$$\frac{1}{n^2}\sum_x \#\{(w,y) \mid w \sim x \sim y\}^2.$$

And we can now do the same thing — we can fold along x, and we get that this is

$$\frac{1}{n^2}\left(\sum_x \deg(x)^2\right)^2 \ge \frac{1}{4}\left(\sum_x \deg(x)\right)^4$$

by Cauchy–Schwarz again. But $\sum_x \deg(x) = 2e(G)$, so we get that this is at least

$$\frac{(2e(G))^4}{n^4} \ge p^4 n^4. \qquad \square$$

Here we've seen an algebraic proof, but it's important to also understand it pictorially (since we might want to do this for much larger graphs). The key point is that Cauchy–Schwarz visually corresponds to reflecting along a symmetry line.

§10.4.2 A map

We'll now prove the equivalences using the following map. We'll first show that Disc and Disc' are equivalent (which won't be too hard). We'll also show

$$\textsc{Disc} \to \textsc{Count} \to C_4 \to \textsc{Codeg} \to \textsc{Disc}.$$

Finally, we'll also show that C_4 and Eig are equivalent.

Proof Disc $\to$ Disc′. This is not hard — we can simply set $Y = X$, and note that the difference between $\binom{|X|}{2}$ and $\frac{|X|^2}{2}$ is more or less negligible (and doesn't contribute to anything). $\square$

Proof Disc′ $\to$ Disc. Here we need an idea called a 'polarization identity' — we have

$$e(X,Y) = e(X \cup Y) + e(X \cap Y) - e(X \setminus Y) - e(Y \setminus X).$$

To see why this is true, imagine we take the adjacency matrix and chop it into three sections (X, Y, and their intersection).

We'll prove *twice* this identity — we want the top-right 2×2 and the bottom-left 2×2, as a combination of principal squares. To get this, you can take the whole square and the middle square, and then you can subtract the top-left and bottom-right squares. Every square appears the same number of times on the left and right.

Now once we have this, if each of the terms on the right-hand side satisfies what we want, then so does the left-hand side. $\square$

Remark 10.11. If we've ever thought about inner products vs. norms in Hilbert spaces, it's the same kind of thing — norms and inner products have an equivalent relationship (if you define a norm, then it automatically defines an inner product via the polarization identity). Here $e(X)$ is like a norm, and $e(X, Y)$ is like an inner product.

Now we'll get to the interesting parts. The fact that Count implies C_4 is automatic, since C_4 is a special case. We'll defer the proof that Disc implies Count to the next chapter. It's a *counting lemma*; we sort of saw it in the previous chapter (we proved a lower bound, but the proof can also be turned into an upper bound). Next chapter we'll see a more streamlined way to do it.

Proof $C_4 \implies$ *Codeg.* Assume C_4. Then we have that

$$\sum_{u,v} \operatorname{codeg}(u,v) = \sum (\deg x)^2 \ge \frac{1}{n} (\sum_x \deg x)^2$$

by Cauchy–Schwarz (as we saw earlier). But we know $e(G) = p\binom{n}{2} + o(n^2)$, so $\sum_x \deg(x) = pn^2 + o(n^2)$, and therefore we have

$$\sum_{u,v} \operatorname{codeg}(u,v) \ge p^2 n^3 + o(n^3).$$

On the other hand, we have

$$\sum_{u,v} \operatorname{codeg}(u,v)^2 = \#\text{labelled } C_4\text{'s} + O(n^3)$$

(the left-hand side counts labelled walks of length 4, and the only difference is repetition of vertices). By assumption the number of C_4's is $p^4 n^4 + o(n^4)$.

So on one hand, we have a lower bound on the mean and an upper bound on the *squared mean*, which we can interpret as an upper bound on the variance. This tells us that the codegrees should all be very concentrated — by Cauchy–Schwarz we get that

$$\frac{1}{n^2} \left(\sum_{u,v} \left| \operatorname{codeg}(u,v) - p^2 n \right| \right)^2 \le \sum_{u,v} (\operatorname{codeg}(u,v) - p^2 n)^2.$$

If you expand the right-hand side you get 3 terms; you upper-bound the first and lower-bound the second, and the third term is a constant. This gives that this quantity is at most

$$p^4 n^4 - 2p^4 n^4 + p^4 n^4 + o(n^4).$$

The main terms all cancel out, and the conclusion is $o(n^4)$.

There is some calculation here, but two things happened. It's not harder than the calculation we did before; and the intuition is that the number of edges gives you a lower bound on the codegree, while the C_4-bound gives you an upper bound on their squares. This means the codegrees have very little variance, so are close to their mean. □

We can think of going from Count to C_4's to Codeg to Disc as 'downstream' (we're going from something more complicated to something less complicated).

Codeg $\implies$ *Disc.* F First we'll show that Codeg implies *concentration* of degrees — that

$$\frac{1}{n} \left(\sum_u |\deg u - pn| \right)^2$$

is small (which intuitively says that the degrees are concentrated around pn). To prove this, we can (again) use Cauchy–Schwarz — this is at most

$$\sum_u (\deg u - pn)^2 = \sum_u (\deg u)^2 - 2pn \sum_u \deg u + p^2 n^3.$$

The first term $\deg^2$ is the same as counting two-edge walks, and so has to do with codegree — so we can rewrite this as

$$\sum_{xy} \operatorname{codeg}(x, y) - 4pne(G) + p^2 n^3.$$

Now we have a codegree condition; and by the triangle inequality, dropping the absolute value signs, we get that this is

$$p^2 n^3 - 2p^2 n^3 + p^2 n^3 + o(n^3).$$

The first three terms cancel out, and we get what we want.

We've now seen that the degrees are concentrated. But this itself is not equivalent to quasirandomness — you can take a regular graph (for example, two disjoint cliques) where every vertex has *exactly* the same degree, but it's not quasirandom. So even though codegree distributions are enough for quasirandomness, degree distribution is not.

But it's still useful; we'll now use it to bound the expression in DISC (combined with the codegree condition).

We wish to bound

$$\frac{1}{n} |e(X, Y) - p |X| |Y||^2 = \frac{1}{n} \left(\sum_{x \in X} \deg(x, Y) - p |Y| \right)^2.$$

By Cauchy–Schwarz, this is at most

$$\sum_{x \in X} (\deg(x, Y) - p |Y|)^2.$$

The next step is a simple observation but core to the proof — this Cauchy–Schwarz step made all the summands nonnegative, and now we are free to add more summands. So now we'll change X to the whole vertex set; and this is at most

$$\sum_{x \in V} (\deg(x, Y) - p |Y|)^2.$$

Then continuing, we can expand as

$$\sum_{x \in V} \deg(x, Y)^2 - 2p |Y| \sum_{x \in V} \deg(x, Y) + p^2 n |Y|^2.$$

The first term is the number of edges from x to Y squared — so we're counting *pairs* of edges from x to Y. We can reverse the order and sum over pairs of vertices in y, and count their codegrees — then we get

$$\sum_{y, y' \in Y} \operatorname{codeg}(y, y') - 2p |Y| \sum_{y \in Y} \deg(y) + p^2 n |Y|^2$$

(the second term counts edges from the entire vertex set to Y; that's the same as summing degrees in Y).

For the first term, were summing codegrees — not *all* the codegrees, but just ones between vertices in Y. But we can simply take CODEG and drop all the irrelevant terms, and we get an estimate (with the appropriate error bound). Likewise, for the second term we can use what we just proved about the degree distribution. So this is at most

$$|Y|^2 p^2 n - 2p |Y| \cdot |Y| pn + p^2 n |Y|^2 + o(n^3).$$

As we would expect, the main terms all cancel and we are left with a sub-cubic error term. □

Remark 10.12. The codegree condition involves a sum over the entire graph; but if we keep only a fraction of the terms, then the sum is still small (since all the terms are positive). (It's crucial that we have the absolute values — otherwise the statement is satisfied by any regular graph.)

We've now completed most of the implications; the only thing left is eigenvalues. For this, we'll need some basic facts about eigenvalues from linear algebra.

Let A be a $n \times n$ real symmetric matrix. (If A is *not* symmetric, then it's very hard to think about eigenvalues. We'll be working with adjacency matrices.) Then by the Spectral theorem, we have real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, and eigenvectors $v_1, \ldots, v_n$ which can be taken to be orthonormal.

Theorem 10.13 (Courant–Fischer min-max Theorem)

We have

$$\lambda_1 = \max_{v \in \mathbb{R}^n \setminus \{0\}} \frac{\langle v, Av \rangle}{\langle v, v \rangle}.$$

Once we fix a choice of the top eigenvector v_1, we have

$$\lambda_2 = \max_{v \perp v_1} \frac{\langle v, Av \rangle}{\langle v, v \rangle}.$$

(This is a version of the first two eigenvalue characterizations; they're important for us and good to know in general.)

We'll also need the following trace formula: we have

$$\operatorname{tr} A^k = \lambda_1^k + \cdots + \lambda_n^k,$$

and $\operatorname{tr} A^k$ counts closed length k walks. You can think of this as a way to link the physical quantity of counting walks to the spectral quantity of eigenvalues.

Proof EIG $\to C_4$. Let A be the adjacency matrix of G. The number of labelled 4-cycles is essentially the number of walks of length 4, up to a small error; so it is $\operatorname{tr} A^4 + O(n^3)$.

Now let's look at $\operatorname{tr} A^4 = \lambda_1^4 + \cdots + \lambda_n^4$. We know that the top eigenvalue is pn, and the others are $o(n)$; this means

$$\operatorname{tr} A^4 = p^4 n^4 + o(n^4) + \sum_{i=2}^{n} \lambda_i^4.$$

The first term is supposed to be the main term, and we want to show all the other terms are small because $|\lambda_i| = o(n)$. But this requires some calculation. There's a subtle but important lesson here — it's easy to do this calculation incorrectly. We'll first do so to show what goes wrong — you might say that $\sum_{i \geq 2} \lambda_i^4 \leq n \cdot \max_i |\lambda_i|^4$. But if we do this, then we get $o(n^5)$, which is too big. (This is correct but unhelpful.)

So there's an important trick we'll see now and later — we don't need to be wasteful in this step, and we can instead leave behind a L^2. We have

$$\sum_{i \geq 2} \lambda_i^4 \leq \max_{i \neq 1} |\lambda_i|^2 \sum_{i=1}^{n} \lambda_i^2.$$

And this L^2 that we left is $\operatorname{tr} A^2 = 2e(G) \leq n^2$. Putting these together, we get that this is $o(n^2) \cdot \operatorname{tr} A^2 = o(n^4)$, which is what we want. □

The only remaining part is to show that C_4 implies EIG. In a way we'll run this proof in reverse, but it helps to first show a lemma:

Lemma 10.14

The top eigenvalue is always at least the average degree.

Proof. Using $v = \mathbf{1}$ to be the vector of all 1's, we have

$$\lambda_1 \geq \frac{\langle \mathbf{1}, A\mathbf{1}\rangle}{\langle \mathbf{1}, \mathbf{1}\rangle} = \frac{2e(G)}{v(G)} = \text{avg deg.}$$ □

Proof $C_4 \implies$ Eig. We have

$$\sum_{i=1}^{n} \lambda_i^4 = \operatorname{tr} A^4 \leq p^4 n^4 + o(n^4)$$

by the assumption on C_4. On the other hand, we also know that $\lambda_1 \geq pn + o(n)$. So we already have a big term on the left that takes up almost all the possible contributions, which implies everything else must be small — we must have $\lambda_1 = pn + o(n)$ and $\lambda_i = o(n)$. □

This finishes the proof of the Chung–Graham–Wilson theorem; we'll now discuss some facts and questions that arose.

§10.5 The role of C_4

In the theorem, C_4 plays a somewhat special role — this is the kind of cleanest and simplest condition, and initially seems the weakest. Can you replace C_4 by something else with the same effect — which graphs F can take the place of C_4?

The triangle doesn't work. One way to see this is that we saw a theorem saying the number of C_4's is *at least* random, while the number of triangles could even be 0 — so it can go in both directions (whereas C_4's can only go up).

The question of which F work is difficult, and is known as the *forcing conjecture*. If F is a tree then it is not forcing (any regular graph will have the same number of trees). But the conjecture is that anything that's bipartite and not a tree is good enough.

Conjecture 10.15 (Forcing conjecture) — A graph is forcing if and only if it is bipartite and not a tree.

This is very open, but we know that complete bipartite graphs and cycles are forcing.

§10.6 Sparse graphs

The next thing we'll address is what happens for sparse graphs. This is a theorem, so it is true for any value of p, and you are allowed to take $p = 0$. But somehow this is not satisfying — if you have a sparse sequence of graphs, then subquadratic error is too much error to give (it hides all the interesting things that could happen if your graph has edge density e.g. $n^{-0.1}$). So the theorem itself becomes uninteresting.

So here are some formulations of sparse versions of quasirandomness properties. We're not stating a theorem, just writing some definitions.

Definition 10.16 (Sparse-DISC). Suppos that $p = p_n \to 0$. Then SparseDISC is the property that $|e(X,Y) - p|X||Y|| = o(pn^2)$.

It'd be meaningless to have $o(n^2)$, since both terms on the left are themselves $o(n^2)$. So we need to take the correct scaling.

Definition 10.17 (SparseCOUNTH). $\#H = (1+o(1))p^{e(H)}n^{v(H)}$.

Definition 10.18 (SparseEIG). $\lambda_1 = (1+o(1))pn$, and $|\lambda_i| = o(pn)$ for all $i \geq 2$.

These are just statements of properties; then you can ask, are they equivalent? The answer is no — they are not equivalent. In particular, SparseDISC does *not* imply SparseCOUNT. Why not? Imagine that $p = n^{-c}$; for simplicity $\frac{1}{2} < c < 1$ (think of c as 0.6). In $G(n,p)$, the number of triangles is on the order of n^3p^3, while the number of edges is typically close to n^2p. When p is quite small, the number of triangles is much smaller than the number of edges.

So what you can do then is remove all the triangles, by removing an edge from each triangle. This ends up removing $o(pn^2)$ edges.

We claim that the resulting graph satisfies SparseDISC — because $G(n,p)$ satisfies it, and we removed so few edges that it doesn't affect the bound. But SparseCOUNT, at least for triangles, is certainly violated since we've gotten rid of all the triangles.

So this is an example where SparseDISC doesn't imply SparseCOUNT.

That's not the end of the story — the last chapter of the book is about how to make this implication work under additional hypotheses. It turns out that the proof of the Green–Tao theorem, about arithmetic progressions in the primes (which are sparse) hinges on understanding how to make this work under additional assumptions. (Just as how we saw Roth's theorem and Szemerédi's theorem needed a counting lemma, here we need a counting lemma in a sparse setting with additional hypotheses.)

There are *some* implications that are fine. Which of these implications should be okay in sparse settings? The DISC vs DISC' is okay (it's just about counting edges). DISC to COUNT is bad. COUNT to C_4 is just a special case. What about C_4 to CODEG? We're applying Cauchy–Schwarz, and this is still fine. Likewise CODEG to DISC is also fine.

The intuition is that hwen you fold the graph to make it smaller, this is okay; but when you want to go upstream (going to more complicated graphs) you start having problems.

WHat about C4 vs. EIG? We had an inequality which was $o(p^2n^2)O(pn^2)$; this is bad because we need p^4n^4. In the second part, it's tricky because we have to be careful with the $o(n^4)$. The trace is a better object to look at; if you want C_4 to really be a statement about the trace, then it should be okay.

§11 October 16, 2023

We started talking about pseudorandom graphs — specifically, last lecture we saw the notion of a *quasirandom graph*. Chung–Graham–Wilson introduced this notion in the 1980s; they showed that there are many equivalent notions of what it means for a graph to be quasirandom (we proved this theorem last lecture).

Remark 11.1. Fan Chung will give a guest lecture next Wednesday.

In the Chung–Graham–Wilson equivalences, there were several things. There was a discrepancy condition, similar to the ε-regular pair condition. There was a C_4 condition, which is surprising because it's just about a single graph count — just having the right number of 4-cycles is enough to force quasirandomness.

There's also a condition about eigenvalues — all the eigenvalues except the top one being small. Today we'll further explore the role of eigenvalues in graph theory.

§11.1 Expander mixing lemma

This is an important theorem about the relationship between mixing and eigenvalues in graphs.

We will work in this lecture with d-regular graphs. (Note that d-regular means every vertex has degree d.) Everything is nicer with d-regular graphs; you can formulate some of this for non-regular graphs as well, but we won't do this.

Definition 11.2. A (n, d, λ)-graph is an n-vertex d-regular graph such that the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ of its adjacency matrix satisfy the property that $\max_{i \neq 1} |\lambda_i| \leq \lambda$.

Remark 11.3. We always have $\lambda_1 = d$, with the eigenvector $\mathbf{1}$.

So a (n, d, λ)-graph is a n-vertex d-regular graph where the top eigenvalue is d, and the other eigenvalues are all small (controlled by λ).

The expander mixing lemma says the following.

Theorem 11.4 (EML)
If G is an (n, d, λ)-graph, then for all $X, Y \subseteq V(G)$, we have

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

Before we prove it, we'll discuss some aspects. The left-hand side is the same as what we saw in the discrepancy condition last lecture. But the right-hand side is much stronger — on the right-hand side we essentially had εn^2, but here we have something potentially quite a bit smaller. Last time, if you made X and Y smaller, you didn't get to decrease the error; but this lemma gives you better control for smaller X and Y. So it's a much stronger statement than what discrepancy buys you.

§11.1.1 Linear algebra

Before we prove the EML, we'll review some linear algebra.

Definition 11.5. For any (not necessarily symmetric) matrix $A \in \mathbb{R}^{m \times n}$, its *operator norm* is

$$\|A\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|Ax|}{|x|}.$$

The operator norm considers A as a map $x \mapsto Ax$, and considers how much it can amplify the length of a vector. We can equivalently define it as

$$\|A\| = \sup_{x, y \neq 0} \frac{\langle y, Ax \rangle}{|x| |y|}.$$

For a real symmetric matrix, the operator norm can be read from its eigenvalues — we have

$$\|A\| = \max_i |\lambda_i(A)|.$$

For general matrices A, we have that $\|A\|$ is the largest *singular value* of A, though we won't use that here.

§11.1.2 Proof of EML

With that in mind, let's prove the expander mixing lemma.

The content of EML is to view the LHS as an expression of the form $\langle y, Ax\rangle$ — x and y will basically be the indicator vectors of X and Y as sets, and for an appropriate matrix A, the desired inequality will be exactly what the definition of operator norm gives us.

Consider the matrix

$$A_G - \frac{d}{n}J$$

where A_G is the adjacency matrix of the graph, and J is the all-ones matrix (so we're subtracting enough to create a matrix with mean 0).

What are the eigenvalues of this matrix? Suppose A_G has eigenvalues $\lambda_1 = d \geq \lambda_2 \geq \cdots$ with eigenvectors $\mathbf{1}$, v_2, v_3, $\ldots$, v_n. What happens here? Well, $\mathbf{1}$ is also an eigenvector of J. So $\mathbf{1}$ remains an eigenvector of $A_G - \frac{d}{n}J$, with eigenvalue 0.

On the orthogonal complement of $\mathbf{1}$, the action of J is just 0. So all the other eigenvectors remain the same, with the same eigenvalues.

So with that in mind, we can apply the (n, d, λ)-hypothesis — since G is (n, d, λ), then

$$\left\| A - \frac{d}{n}J \right\| \leq \lambda.$$

Now we can rewrite $e(X, Y) - \frac{d}{n}|X||Y|$ in terms of matrices as

$$e(X, Y) - \frac{d}{n}|X||Y| = \langle \mathbf{1}_X, \left(A_G - \frac{d}{n}J\right)\mathbf{1}_Y\rangle.$$

Now applying the spectral norm condition, we have

$$\left| e(X, Y) - \frac{d}{n}|X||Y| \right| = \left| \langle \mathbf{1}_X, \left(A_G - \frac{d}{n}J\right)\mathbf{1}_Y\rangle \right| |\mathbf{1}_X||\mathbf{1}_Y|.$$

The first term is at most λ; the second is $\sqrt{|X|}$, and the third is $\sqrt{|Y|}$. This gives the desired inequality.

The EML tells you that if all the eigenvalues except the first one (which has to be large) are small, then your graph has very good discrepancy properties. In fact this is stronger than the discrepancy condition we saw last time, because this quantity is still meaningful for sparser graphs — the right-hand side is still meaningful for smaller sets X and Y. Last time we saw that for dense graphs, having small second eigenvalue is equivalent to the other conditions, but this isn't true for sparse graphs; but this stronger statement can be applied.

For pseudorandom graphs, there's a similar story — what happens for very sparse graphs? This goes under the name of expanders; these are related, but not exactly the same.

§11.2 Expanders

For this, think about sequences of graphs with fixed degree (e.g. a sequence of 3-regular or 10-regular graphs).

A concept you'll often hear in discussions of spectral graph theory (or computer science) is the *spectral gap* — this is defined as $\lambda_1 - \lambda_2$. This is a very important quantity, for the following reason. It has a relationship to the *edge-expansion ratio*

$$h(G) = \min_{\substack{S \subseteq V \\ 0 < |S| \leq \frac{1}{2}|V|}} \frac{e_G(S, V \setminus S)}{|S|}.$$

What does it mean to have large edge-expansion ratio? It means that in the graph G, whenever I take a vertex set that is relatively small (no more than half the vertex sets) and count how many edges *leave* this set per vertex in S (on average), this number is always bounded below by the edge-expansion ratio. So if some set has edge-expansion ratio at least 1, this means every set has at least as many edges leaving it as vertices inside.

This is a property of graphs that holds for some graphs but not others. Roughly speaking, graphs with 'large' edge-expansion ratios are called 'good expanders.' (We say roughly because in specific applications you may want more precise notions.)

In a random graph, somehow at every vertex the growth is tree-like — so we have very rapid expansion. That's an example of a good expander.

Here are some examples of not good expanders. One is a graph that looks like a dumbbell — maybe there's a lot of interesting things happening in one half, a lot of interesting things happening in the other, and some very thin paths connecting them. Then we can take one half; it has lots of vertices but very few edges leaving it.

Another bad expander is the integer grid. If you take an $\ell \times \ell$ box of the grid, then you have ℓ^2 vertices but only 4ℓ edges leaving it; and this ratio goes to 0 as ℓ gets large. So the graph doesn't expand — it grows very slowly. So expansion measures how fast sets grow as you expand them.

These are very important properties; they come up in lots of settings. In discrete mathematics, they come up because graphs with good expansion properties are useful in computer science — they also have very good mixing behavior (a random walk tends to mix very quickly). If you want to use a random walk to get pseudorandom behavior, or for sampling applications, then you want something that mixes quickly; these graphs are particularly suited for those purposes.

Another setting expansion comes up is geometry. Imagine you have some space (e.g. a manifold); does it look like a dumbbell (where you can squeeze it), or something much more rapidly expanding?

§11.3 Cheeger's inequality

Some of the earliest explorations of this subject came from Riemannian geometry; Cheeger's inequality in particular began in geometry (Riemannian manifolds), and was transfered to graphs by Alon–Millman.

Theorem 11.6 (Cheeger's inequality)

If G is a n-vertex d-regular graph whose adjacency matrix has spectral gap $\kappa = d - \lambda_2$, then its edge-expansion ratio $h = h(G)$ satisfies

$$\frac{\kappa}{2} \le h \le \sqrt{2d\kappa}.$$

In particular, for a sequence of d-regular graphs for fixed d, having a lower bound on the edge-expansion ratio — i.e., knowing that $\inf h(G_n) > 0$ — is equivalent to having a lower bound on the spectral gap — i.e., $\inf \kappa(G_n) > 0$. We call a sequence of graphs (G_n) that satisfies this property *expanders*.

Informally, graphs that look like rapidly-growing trees (like the random graph) are expanders; graphs that look like the grid are not.

Example 11.7

Consider the graphs $\{0,1\}^d$ (the Boolean cube — note that these don't exactly have bounded degree). Here the eigenvalue distribution can be computed exactly; the eigenvalues are $-d$, $-d+2$, $\ldots$, $d-2$, d, and their multiplicities are exactly the binomial coefficients. In particular, $\kappa = 2$.

Meanwhile, what's the worst set in terms of the number of edges going out of it? Take the bottom half of the cube; then we get exactly one edge per vertex. It turns out that this is the worst case, so $h = 1$; this shows the tightness of the lower bound.

Example 11.8

Consider a cycle C_n. Later in this lecture we'll compute its eigenvalues; the top eigenvalue is 2, and the next is $2\cos\frac{2\pi}{n}$; so $\kappa = 2 - 2\cos\frac{2\pi}{n} = \Theta(\frac{1}{n^2})$.

Meanwhile, we can take a set consisting of half the cycle; then we only have 2 edges going out, so $h = \frac{2}{\lfloor n/2 \rfloor} = \Theta(\frac{1}{n})$.

Proving Cheeger's inequality will be on the homework (we'll be guided through it).

§11.4 Cayley graphs

An important class of graphs that come up are Cayley graphs — graphs that are built from groups.

Definition 11.9. Let Γ be a finite (usually abelian) group, and let $S \subseteq \Gamma$ be a generating set which is closed under inverses (i.e., $s \in S$ if and only if $s^{-1} \in S$). We define the Cayley graph $\mathrm{Cay}(\Gamma, S)$ to be the graph on vertex set Γ, where the edges are of the form $g \sim gs$ whenever $g \in \Gamma$ and $s \in S$.

For non-abelian groups, we need to be consistent as to whether we write s on the right or left; but as long as we do it consistently, it doesn't matter which way you do it.

We'll be particularly interested in the case of $\Gamma = \mathbb{Z}/p\mathbb{Z}$, but other abelian groups are in a way quite similar (you can write each as a direct product of similar objects).

Example 11.10

We have $\mathrm{Cay}(\mathbb{Z}/n\mathbb{Z}, \{\pm 1\}) \cong C_n$ (where C_n is the n-vertex cycle).

A particularly important example of a Cayley graph is the Paley graph, the Cayley graph corresponding to quadratic residues.

Definition 11.11. Let $p \equiv 1 \pmod 4$ be a prime. The Paley graph is $\mathrm{Cay}(\mathbb{Z}/p\mathbb{Z}, S)$ where $S = \{x^2 \mid x \in (\mathbb{Z}/p\mathbb{Z})^*\}$ is the set of nonzero quadratic residues.

The reason we need $p \equiv 1 \pmod 4$ is that -1 is a quadratic residue if and only if $p \equiv 1 \pmod 4$.

We mentioned this graph last time as an example of a quasirandom graph; today we'll prove this. (We'll develop some techniques to show the Paley graph is indeed quasirandom.) For this, we'll need to say what are the eigenvalues of the Paley graph (we'll prove quasirandomness through considering its eigenvalues).

More generally, there is a way to compute the eigenvalues of any Cayley graph on a cyclic abelian group. The general message here is that the eigenvalues of $\mathrm{Cay}(\mathbb{Z}/p\mathbb{Z}, S)$ (where S can be anything) are in complete correspondence with the Fourier transform of 1_S. (This is a general principle, and it's also a link between the world of graph theory and additive combinatorics — we will see the Fourier transform plays a large role in additive combinatorics, and eigenvalues link these concepts together.)

§11.5 Eigenvalues of Cayley graphs

Question 11.12. What are the eigenvalues and eigenvectors of $\mathrm{Cay}(\mathbb{Z}/n\mathbb{Z}, S)$ (where S is an arbitrary set)?

(Here n doesn't have to be a prime.)

First, we'll say what the *eigenvectors* are. It turns out that the eigenvectors do not depend on S (this can be viewed as a statement about 'simultaneous diagonalization').

We'll have vectors of length n. The first eigenvector v_0 is the all-1's vector; because we want our eigenvectors to have length 1, we renormalize to

$$\sqrt{n}v_0 = (1, 1, \ldots, 1).$$

The next eigenvector is going to be as follows. Let $\omega = e^{2\pi i/n}$ be an nth root of unity. Then

$$\sqrt{n}v_1 = (1, \omega, \omega^2, \omega^3, \ldots, \omega^{n-1}).$$

Similarly, we have

$$\sqrt{n}v_2 = (1, \omega^2, \omega^4, \omega^6, \ldots, \omega^{2(n-1)}),$$

and so on, up to

$$\sqrt{n}v_{n-1} = (1, \omega^{n-1}, \omega^{2(n-1)}, \ldots).$$

Note that these vectors do not depend on S — they only depend on n.

We need to prove that these are eigenvectors. Let's see what happens if we multiply these vectors by the adjacency matrix of our graph — let A_G be the adjacency matrix, and consider $\sqrt{n}A_G v_j$. The coordinate at location $x \in \mathbb{Z}/n\mathbb{Z}$ is equal to

$$\sum_{s\in S} \omega^{j(x+s)}.$$

(We have a matrix A and a vector v_j; imagine that we take the xth row of A and perform this scalar multiplication. The xth row has a bunch of 1's at columns $x + s$ for $s \in S$; we add those entries together and get this.) But we can rewrite this sum as

$$\left(\sum_{s\in S} \omega^{js}\right) \omega^{jx}.$$

Let $\lambda_j = \sum_{s\in S} \omega^{js}$; note that λ_j does not depend on x. So then we see that

$$Av_j = \lambda_j v_j$$

(since this is true at all the coordinates, it's true as a vector equality).

So to summarize, the eigenvalues of A_G are given by

$$\lambda_j = \sum_{s\in S} \omega^{js}$$

as j goes from 0 to $n - 1$. Unlike the convention we used before, these are not necessarily ordered (when $j = 0$, we do have $\lambda_0 = |S|$; but the rest are unsorted).

So we have a formula that gives the eigenvalues as a function of S, which is quite simple. We'll see this come up again later in the course, in that this is also the formula for the Fourier transform. And the $n \times n$ matrix consisting of the v_j's is referred to as the discrete Fourier transform matrix; it's an important matrix. (If you input a vector, then it outputs the Fourier transform.)

§11.6 Eigenvalues of the Paley graph

This is a formula for the eigenvalues of any Cayley graph on $\mathbb{Z}/n\mathbb{Z}$; let's apply it to our specific setting of the Paley graph, where S is the set of quadratic residues.

Theorem 11.13

Let $p \equiv 1 \pmod 4$. Then the eigenvalues of the Paley graph (for p) are $(p-1)/2$ once, then $(\sqrt{p}-1)/2$ repeated $(p-1)/2$ times, and $(-\sqrt{p}-1)/2$ repeated $(p-1)/2$ times.

Proof. We have that

$$\lambda_j = \sum_{s \in S} \omega^{js} = \frac{1}{2}\left(-1 + \sum_{x \in \mathbb{Z}/p\mathbb{Z}} \omega^{jx^2}\right)$$

(since the sum counts everything twice and has an $x = 0$ term). So we essentially want to show that $\sum_{x \in \mathbb{Z}/p\mathbb{Z}} \omega^{jx^2}$ is $\sqrt{p}$ in absolute value. This is an important fact known as the Gauss sum:

Theorem 11.14 (Gauss)

We have

$$\left|\sum_{x \in \mathbb{Z}/p\mathbb{Z}} \omega^{jx^2}\right| = \sqrt{p}.$$

Once we have the Gauss sum, the theorem follows, though we have to say a bit more. First, we know the eigenvalues of the graph are real (since we have a symmetric real matrix); so we know that we have some number of eigenvalues $(\sqrt{p}-1)/2$ and some $(-\sqrt{p}-1)/2$. But we also know that the sum of eigenvalues is the trace of the matrix, which is 0; and the only way this can occur is for them to be split equally.

Now let's talk about the Gauss sum. This is an important computation, based on a simple but clever idea — to evaluate this quantity, we square the left-hand side and manipulate the expression using change of summation.

Proof. We have

$$\left|\sum_{x \in \mathbb{Z}/p\mathbb{Z}} \omega^{jx^2}\right|^2 = \sum_{x,y \in \mathbb{Z}/p\mathbb{Z}} \omega^{jx^2 - jy^2}$$

(the $-$ sign comes from complex conjugation). But we can relabel variables to rewrite the sum as

$$\sum_{x,y \in \mathbb{Z}/p\mathbb{Z}} \omega^{j(x+y)^2 - jy^2} = \sum_{x,y \in \mathbb{Z}/p\mathbb{Z}} \omega^{j(2xy+y^2)}.$$

For each fixed y, we can figure this out, since we just have a simple exponential sum — if $y \neq 0$ then we're summing roots of unity in some order, so the sum is 0. If $y = 0$, then we have a sum of p numbers all equal to 1 (as the exponent is always 0), so it's equal to p. This means the sum is p, as desired. □

□

§11.7 Non-abelian groups

In a way, this is the story for all abelian groups; we just talked about $\mathbb{Z}/n\mathbb{Z}$, but any finite abelian group can be written as a direct sum, and you can also define the notion of a Fourier transform (there'll be more factors, but it's the same story).

Non-abelian groups are very different. For the Cayley graphs of abelian groups, quasirandomness depends on the generators. For example, if we took $S = [-\frac{p}{4}, \frac{p}{4}] \setminus \{0\}$, then the graph would look like a circle where each vertex would be connected to the close-by vertices; this is not quasirandom (consider a small patch of vertices on the top and bottom; there are no edges between them). On the other hand, Paley graphs are quasirandom. So depending on which generators you use, you may or may not get quasirandom Cayley graphs.

For *certain* nonabelian groups, *all* Cayley graphs are quasirandom. So quasirandomness is a property of the group, rather than that of the generating set. What property is required there? This leads to the notion of *quasirandom groups*.

Definition 11.15 (Quasirandom groups)**.** A group is *quasirandom* if it has no small nontrivial representations.

We'll be somewhat sketchy in this part of the lecture, but given a group, there's a standard theory of group representations; and there's such a thing as the smallest nontrivial representation, which has a size; that size is either small or large.

For abelian groups, *all* irreducible representations are 1-dimensional. So abelian groups never have this property. But some non-abelian groups do have this property. And Gowers showed that for groups with these properties, all Cayley graphs are quasirandom; and the converse is true as well.

Let's explain why Gowers was looking at this problem. It came from the following natural combinatorial question:

Question 11.16. Given a group of order n, what is the size of its largest product-free subset?

Example 11.17

In $\mathbb{Z}/n\mathbb{Z}$ (where product-free means sum-free), we can have a set that is linear in size but sum-free — for example, $[\frac{n}{3}, \frac{2n}{3}]$ has no solutions to $x + y = z$ and has linear size.

Question 11.18. Can you always find a linear-sized product-free set?

What Gowers showed is that for quasirandom groups the answer is no.

This connects two of the topics we saw in today's lecture, and we'll now explain the connection. We need a couple of facts.

First, what's the relationship between quasirandom groups (in terms of irreducible representations) and eigenvalues?

Theorem 11.19

If Γ is a group of order n with no nontrivial representations of dimension less than k, then every d-regular Cayley graph on Γ is an (n, d, λ)-graph with $\lambda < \sqrt{dn/k}$.

The trivial bound on λ is d — λ is the second-largest eigenvalue in absolute value. And you can think of d as being on the same order as n (for dense graphs). If n grows asymptotically, then this quantity is $o(n)$, which is what we need for Chung–Graham–Wilson quasirandomness.

Proof. The idea is that given an eigenvector v that is nontrivial (orthogonal to $\mathbf{1}$), we can have Γ act on v to get at least k eigenvectors with the same eigenvalue (so the eigenspace has dimension at least k). Then we have

$$\operatorname{tr} A^2 = \sum \lambda_i^2 \geq d^2 + k\mu^2,$$

where $\mu = \max_{i \geq 2} |\lambda_i|$ is the second-largest eigenvalue in absolute value. On the other hand, $\operatorname{tr} A^2$ is theh sum of entries of A, which is nd. So $\mu \leq \sqrt{nd/k}$. □

The relevance to the problem of product-free sets lies in a connection relating to what we did in the first part of the lecture, namely the expander mixing lemma.

Theorem 11.20 (Mixing in quasirandom groups)

Let Γ be a finite group with no nontrivial representations of dimension less than k. Let $X, Y, Z \subseteq \Gamma$. Then

$$\left| \#\{(x,y,z) \in X \times Y \times Z \mid xy = z\} - \frac{|X|\,|Y|\,|Z|}{|\Gamma|} \right| < \sqrt{\frac{|X|\,|Y|\,|Z|}{k}}.$$

So we're counting solutions to $xy = z$. We compare to what you'd expect if everything acted perfectly randomly — we take x arbitrarily and y arbitrarily, and we'd expect the probability xy lies in Z (based on density considreations) to be $|Z|/|\Gamma|$. The theorem states that the actual count is 'close' to this naive expectation.

The left-hand side is of quadratic order (a priori). For the right-hand side, if k is increasing then it's subquadratic; so this gives a good bound for large values of k.

This is sort of a special case of the expander mixing lemma (we need a bipartite version, but the same is true there); they look similar (we again have $\sqrt{dn/k}$), and if we put things together correctly then the right inequality pops up.

Proof. Consider the following graph: we consider a bipartite graph where our two sets of vertices are both Γ, and we draw edges xz (with x on the left and z on the right) if $y = x^{-1}z \in Y$. This is called the *bipartite Cayley graph* $\operatorname{BiCay}(\Lambda, Y)$.

Consider some subset X on the left and Z on the right. By a bipartite generalization of the expander mixing lemma (which we haven't stated, but it's the exact same proof; instead of having one graph, we have a bipartite graph and consider the number of edges between a part on the left and a part on the right), we see that this graph is a bipartite (n, d, λ)-graph with $n = |\Lambda|$, $d = |Y|$, and $\lambda < \sqrt{|\Gamma|\,|Y|/k}$ (this fact comes from the previous theorem). Then plugging the various quantities into the EML gives that the number of edges between X and Y — which is the number of solutions — compared to the relevant quantity is bounded (by the quantity coming out of EML). □

§11.7.1 Examples of quasirandom groups

We've talked about quasirandom groups in abstract, but haven't seen examples; we'll now do so.

Example 11.21

The group $\mathbb{PSL}(2,p)$ consists of 2×2 matrices with determinant 1, where we consider A and $-A$ as equivalent. Frobenius showed that all nontrivial representations have dimension at least $(p-1)/2$.

In a way, this is the best known example of a quasirandom group in terms of quantitative bounds (an order-n group whose nontrivial representations are as large as possible); although no one has proven that this is indeed optimal.

Corollary 11.22

Every product-free subset of $\mathbb{PSL}(2,p)$ has size $O(p^{3-1/3})$.

The reason we write it this way is that $|\mathbb{PSL}(2,p)| = \Theta(p^3)$; so every product-free subset has size significantly smaller than the order of the group (though we don't know the actual size of the largest product-free subsets).

Another example that behaves worse quantitatively, but for which we have much better understanding, is the alternating group:

Example 11.23

The *alternating group* A_n (consisting of even permutations) has order $n = m!/2$. Its representations are well-understood (representation theory of symmetric groups is related to Young diagrams and Young tableaux); the smallest nontrivial representation has dimension $m - 1 = \Theta(\log n / \log\log n)$, which increases with the size of the group (though at a logarithmic instead of polynomial rate).

We can again ask for the size of the largest product-free subset; and we now know the precise answer up to a constant factor.

We also know that every simple group other than cyclic groups (which are abelian) is quasirandom.

§11.8 Summary

We saw EML, which connects having small eigenvalues to a mixing property; this is very useful. We also saw an analysis of the spectrum of Cayley graphs, specifically abelian ones; we can compute it quite explicitly, and there's a connection between this object and Fourier transforms (which we'll see later). Using that, we showed the Paley graph is quasirandom (by computing its spectrum).

For nonabelian groups, the situation is interesting in a different way — properties of the group start to matter. For abelian groups, Cayley graphs can be quasirandom or not depending on the generators; for quasirandom groups, all Cayley graphs are quasirandom. An example is $\mathbb{PSL}(2,p)$; one motivating question was showing that $\mathbb{PSL}(2,p)$ (as well as all other quasirandom groups) has no large product-free subsets, which we established by combining the EML and estimates on the eigenvalues.

§12 October 18, 2023

§12.1 Sparse quasirandom graphs

When we talked about quasirandom graphs, we focused on *dense* graphs. We mentioned at the end that you can extend this notion to *sparse* graphs. Here are two relevant definitions. For the purpose of this lecture we'll only discuss d-regular graphs; think of d as sublinear in n.

Definition 12.1. A graph G satisfies sparseDISC(ε) if

$$\left| e(X,Y) - \frac{d}{n}|X||Y| \right| \le \varepsilon dn$$

for all $X, Y \subseteq V(G)$.

Note that here εnd, as opposed to εn^2, is the right order of magnitude.

Definition 12.2. We say G satisfies SparseEIG(ε) if G is an (n, d, λ)-graph with $\lambda \le \varepsilon d$.

This means other than the top eigenvalue, all the other eignevalues are at most λ in absolute value — $\max_{i>1} |\lambda_i| \le \lambda$.

We saw that when d grows linearly with n, these two notions are equivalent; but they are not always equivalent for sparse settings. Today we'll explore that setting further.

Some implications do still hold.

Proposition 12.3

SparseEIG(ε) implies SparseDISC(ε).

Proof. The proof is by an application of the expander mixing lemma, which tells us that

$$\left| e(X,Y) - \frac{d}{n} |X| \, |Y| \right| \le \lambda \sqrt{|X| \cdot |Y|} \le \varepsilon d n$$

since $\lambda \le \varepsilon d$ and $|X|, |Y| \le n$. □

But the converse is not true, even if we loosen the constants — it is *not* true that SparseDISC($o(1)$) implies SparseEIG($o(1)$). Here's an example — take a large random d-regular graph, and add to it a clique K_{d+1} (a very small d-regular graph).

The original graph itself satisfies both properties (we didn't prove this, but it's as random as it gets so it satisfies all the nice properties); think of d as slowly increasing (not as a constant).

When we add K_{d+1}, adding a small number of perturbations doesn't affect discrepancy. But it does affect the eigenvalues — because this creates another eigenvector (you can have the all-ones eigenvector on the left and the all-ones eigenvector on the right; so the top eigenvalue d is repeated twice). So you can easily perturb the eigenvalue property just by local perturbations.

It turns out — quite surprisingly — that among Cayley graphs, the reverse implication *does* hold:

Theorem 12.4 (Conlon–Zhao)

For Cayley graphs, SparseDISC(ε) implies SparseEIG(8ε).

We lose a bit from ε to 8ε, but this doesn't really matter.

Remark 12.5. This is true for any vertex-transitive graph (which rules out the small perturbation construction).

We won't see the full proof (because we'll also talk about other important topics), but we'll see an important tool which was used (which Prof. Zhao learned about when working on this problem) — a fundamental theorem in functional analysis known as Grothendieck's inequality.

§12.1.1 Semidefinite relaxation

The idea is that the left-hand side is about discrepancy (this will come up again next lecture in the language about cut norms), and one way to think about this quantity $e(X,Y) - \frac{d}{n}|X|\,|Y|$ is that it's an expression of the type

$$\sup_{i,j} \sum a_{ij} x_i y_j = x^\intercal \left(A - \frac{d}{n} J \right) y.$$

We want to make this expression as large in absolute value as possible; and we need to select $x_i \in \{0,1\}$ and $y_j \in \{0,1\}$ (since they're supposed to be indicator vectors for the set X and Y).

This quantity is very difficult to compute — if I give you a very large matrix, computing this quantity is a NP-hard problem. But there are ways to go around this — you can try to approximate. In a way we saw this two lectures ago — if you didn't know how to compute this, you could look at 4-cycles. But here we'll look at a different viewpoint that works well in sparse settings, known as *semidefinite relaxation*.

First, we can change the allowed values to $\{\pm 1\}$ (this is just a cosmetic change; you can bound the effect of this by a factor of 4, just by splitting all the ± 1's into $+1$ and -1). The idea is that instead of working with ± 1 as scalars in $\mathbb{R}$, we can work with *unit vectors* in some higher dimensional space $\mathbb{R}^d$. This is a relaxation — we're allowing more possibilities.

But it turns out that by doing this, you make the problem easier to compute (and therefore easier to handle).

Remark 12.6. This is very useful in computer science — for example, computing MAX-CUT of a graph is hard, but using this relaxation together with a randomized rounding step, you can get an approximation (which is the best known one).

We'll do this here. We consider the quantity

$$\sup_{|x_i|\le 1, |y_j|\le 1, x_i, y_j \in \mathbb{R}^d} \sum_{i,j} a_{ij} \langle x_i, y_j \rangle$$

where the supremum is over all *vectors* x_i and y_j in some high-dimensional space with magnitude 1.

What's the relationship between these two quantities? Of course, the first is smaller than the second — we're just allowing more possibilities (and taking the sup). But potentially, the second could be much bigger; this would not be great for us. But it turns out that it cannot be much bigger.

Theorem 12.7 (Grothendieck's inequality)

There exists $K \le 2$ such that

$$K \sup_{x_i \in \{\pm 1\}, y_i \in \{\pm 1\}} \sum_{i,j} a_{ij} x_i y_j \ge \sup_{|x_i|,|y_j|\le 1, |x_i|,|y_j| \in \mathbb{R}^d} \sum_{i,j} a_{ij} \langle x_i, y_j \rangle.$$

(Here d can be anything.)

Grothendieck did very foundational work in functional analysis, before he moved on to reinventing algebraic geometry. This statement says that the semidefinite relaxation is no more than the original problem up to some constant K, known as Grothendieck's constant; and K is at most 2. So by losing a factor of at most 2, we can relax the original problem to the semidefinite relaxation.

Remark 12.8. What's the optimal constant K? This is a fundamental constant that we don't know (though we have some bounds); but we know it's at most 2.

Remark 12.9. If you want to algorithmically estimate the left-hand side, you can do semidefinite programming to compute the right-hand side.

The way the theorem about Cayley graphs is proved is — suppose that you have SparseDISC, and you want to prove that the graph has small eigenvalues. If you have small discrepancy, then the left-hand side is small; so by Grothendieck the right-hand side is small as well. It may not be obvious looking at this expression, but because of the group symmetry — looking at a space $\mathbb{R}^G$ indexed by the elements of the group — the eigenvalues can be expressed in the form on the right-hand side, and this gives a bound on the eigenvalues.

(There's some calculations involved showing that if you have an eigenvector, then the eigenvalues can be expressed in this way. This uses the group symmetry in the graph to encode a lot of the graph as part of the inner product.)

§12.2 Second eigenvalue bounds

Now we'll turn our attention to (n, d, λ)-graphs with fixed d.

Question 12.10. Fix d. What is the smallest possible λ such that there exist infinitely many $(n, d, \lambda + o(1))$-graphs?

(Here $o(1)$ is as $n \to \infty$.)

The answer is $2\sqrt{d-1}$; in the rest of the lecture we'll explain where this comes from. (There are also lots of very important open problems around this problem.)

Alon–Boppana showed the following second eigenvalue bound.

Theorem 12.11 (Alon–Boppana)

Fix d, and let G be an n-vertex d-regular graph whose adjacency matrix eigenvalues are $\lambda_1 \ge \cdots \ge \lambda_n$. Then $\lambda_2 \ge 2\sqrt{d-1} - o(1)$.

This is a lower bound on the second eigenvalue. In other words, if $\lambda < 2\sqrt{d-1}$, then $(n, d, \lambda + o(1))$-graphs don't exist for sufficiently large n.

Remark 12.12. This bound is tight; this is a hard result, which we'll briefly discuss.

This is slightly stronger than the following statement:

Corollary 12.13

For all d and all $\lambda < 2\sqrt{d-1}$, there are only finitely many (n, d, λ)-graphs.

The reason the corollary is slightly weaker is that the Alon–Boppana bound is really about the second-largest eigenvalue, while the (n, d, λ)-graph definition also bounds eigenvalues from below.

We'll now give two proofs of the Alon–Boppana bound. The second proof will only prove the corollary, but they're both interesting and show different ways to think about spectral theory. The two methods that come up most often in spectral graph theory, which are both important, are as follows:

- Examine test vectors, which are closely related to eigenvectors — and try to understand how they behave with respect to the operator (the adjacency matrix).
- Look at the trace of a power of A. Here we don't look at any eigevectors or test vectors; instead we compute the trace of a power of the matrix. This is also called the moment method, and corresponds to analyzing closed walks.

§12.2.1 First proof (Nilli 1991)

Remark 12.14. A. Nilli is a pseudonym used by Noga Alon. It's also the case that Nilli is the name of his daughter. In the book *Proofs from the Book*, there's a chapter that contains this proof; in it you see a picture of Noga's daughter, who is five years old.

For the first proof, we'll need the following lemma, about the construction of a test vector.

Lemma 12.15

Let $G = (V, E)$ be a d-regular graph with adjacency matrix A, and let $r \in \mathbb{N}$ and $st \in E$. Let V_i be the set of all vertices at distance exactly i from st. Define $x = (x_v) \in \mathbb{R}^V$ by

$$x_v = \begin{cases} (d-1)^{-i/2} & \text{if } v \in V_i \text{ for some } i \le r \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\frac{\langle x, Ax\rangle}{\langle x, x\rangle} \ge 2\sqrt{d-1}\left(1 - \frac{1}{r+1}\right).$$

Note that in particular $V_0 = \{s, t\}$; then we walk outwards, and at each step we record V_1 (at distance 1), V_2 (at distance 2), and so on.

In pictures, we have an edge st. The graph then grows in some way starting from st; the top level is V_0, the next level is V_1, the level after is V_3, and so on. The entries of x are 1 in V_0, then $(d-1)^{-1/2}$ in V_1, then $(d-1)^{-1}$ in V_2, and so on.

So we've used this formula to explicitly construct a vector x; and we claim that the quantity $\langle x, Ax\rangle$ has a lower bound. This becomes closer and closer to $2\sqrt{d-1}$ as r gets large.

In a way, this is a straightforward lemma in the sense that it's an inequality with everything defined; but the calculations can potentially be a mess. We'll now see a way to do them nicely.

Proof. Instead of considering the adjacency matrix, we'll consider the Laplacian matrix $L = dI - A$; this has the nice property that

$$\langle x, Lx\rangle = \sum_{uv \in E} (x_u - x_v)^2.$$

(If you expand the right-hand side, it agrees with the definition of the Laplacian.) We also see that trying to compute the numerator $\langle x, Ax\rangle$ is equivalent to computing that of $\langle x, Lx\rangle$. And the nice thing here is that we don't have to worry about interactions within a layer — vertices within a layer have the same x's, so their terms don't matter. So we only have to worry about edges between layers, which saves a lot of computation.

So we can now write

$$\langle x, Lx\rangle = \sum_{i=0}^{r-1} e(V_i, V_{i+1}) \left(\frac{1}{(d-1)^{i/2}} - \frac{1}{(d-1)^{(i+1)/2}}\right)^2 + \frac{e(V_r, V_{r+1})}{(d-1)^r}.$$

(the first term represents all differences between layers up to the final layer; for the final layer we do something slightly different because all the following coordinates are 0).

Note that $e(V_i, V_{i+1}) \le (d-1)|V_i|$ (since we have at most $d-1$ edges coming out — some could go horizontally, but there's always at least one edge going up). In particular, this formula works for the first layer as well (this is why we need to choose an edge st to start with, and not just a single vertex).

With that, we can bound this quantity by

$$\sum_{i=0}^{r-1} |V_i|(d-1)\left(\frac{1}{(d-1)^{i/2}} - \frac{1}{(d-1)^{(i+1)/2}}\right)^2 + \frac{|V_r|(d-1)}{(d-1)^r}.$$

Then we simplify the mess in the parentheses to

$$(\sqrt{d-1} - 1)^2 \sum_{i=0}^{r-1} \frac{|V_i|}{(d-1)^i} + \frac{|V_r|(d-1)}{(d-1)^r}.$$

Finally, we'll absorb some of the last term into the summation — the leading factor expands to $d-2\sqrt{d-1}$, and if we also pull the r term into the summation then we get that this equals

$$(d-2\sqrt{d-1})\sum_{i=0}^{r}\frac{|V_i|}{(d-1)^i}+(2\sqrt{d-1}-1)\frac{|V_r|}{(d-1)^r}.$$

If there were no horizontal edges we'd know exactly the number of edges between layers; but in general we can get an inequality — since $|V_{i+1}|\le(d-1)\,|V_i|$, we know that

$$\frac{|V_r|}{(d-1)^r}\le\frac{|V_i|}{(d-1)^i}.$$

THis means we can bound the above by

$$\left(d-2\sqrt{d-1}+\frac{2\sqrt{d-1}-1}{r+1}\right)\sum_{i=0}^{r}\frac{|V_i|}{(d-1)^i}$$

(where we fully absorbed the last term into the sum). But we have

$$\sum_{i=0}^{r}\frac{|V_i|}{(d-1)^i}=\langle x,x\rangle.$$

So now we're basically done — this implies that

$$\frac{\langle x,Ax\rangle}{\langle x,x\rangle}=d-\frac{\langle x,Lx\rangle}{\langle x,x\rangle}\ge2\sqrt{d-1}-\frac{2\sqrt{d-1}-1}{r+1}\ge\left(1-\frac{1}{r+1}\right)2\sqrt{d-1}.$$ □

This is fairly straightforward once set up nicely, but it's worth asking, why were these the numbers that were chosen? Let's not look at the computation for a second, and step back — suppose you didn't see this proof, but you knew such a proof existed. How would you recover it — why would you choose this vector?

Imagine I give you a graph, and we want to make the quantity

$$\frac{\langle x,Ax\rangle}{\langle x,x\rangle}$$

as large as possible. Abstractly, given a matrix, you'd want to pick the top eigenvector of A. The top eigenvector of A is the all-1's vector; that will not be very good for us for reasons we'll see later on, and it's a boring vector. So let's think beyond the first vector.

In some sense, the critical graph — the graph that plays a critical role in this problem — is the d-regular tree (which we can draw as rooted at an edge). Imagine we have an infinite tree; and we want to understand, what is the top eigenvector on this tree? Then the answer can't be the all-1's vector, because that vector doesn't have bounded L^2 norm. But it's going to be an eigenvector, so we can try to construct this eigenvector. And an eigenvector (with eigenvalue μ) has the property that $Ax=\mu x$. So we can ask, what number should we put on each layer to satisfy this equation? And then we can work in reverse. Imagine that we want to put a different number at each layer, with x_0, x_1, x_2, …; then the equation we need to satisfy is that at every level we have

$$\mu x_i=x_{i-1}+(d-1)x_{i+1}.$$

This is some linear recursion that we need to solve. And the point is that these numbers $(d-1)^{-i/2}$ solve the linear recursion. (This also helps you figure out what is the right value for μ — you can guess the eigenvector from the usual eigenvalue-eigenvector equation, and then use that to come up with the 'pseudo-eigenvector' that we truncate to get a genuine test vector; and that's what feeds into the calculation.)

Remark 12.16. We should think of this test vector as something that came from an eigenvector.

We're not done yet — so far, we just produced a test vector, but we still need to produce a lower bound on λ_2. There's a final step, which we'll do now.

Proof of Alon–Boppana bound. We know that the top eigenvector is $\mathbf{1}$, with eigenvalue d. And from the variational Courant–Fischer characterization of eigenvalues, it suffices to exhibit a vector $z \perp \mathbf{1}$ such that

$$\frac{\langle z, Az\rangle}{\langle z, z\rangle} \geq 2\sqrt{d-1} - o(1)$$

(this would show the second-largest eigenvalue is large). This is kind of what we did — but we're not quite done yet. Why does the x that we produced not satisfy what we want? All its entries are nonnegative, so it's not orthogonal to $\mathbf{1}$. So we need to do something else.

We took a test vector x living in some ball of radius r. But we have a big graph; so we can take another test vector y, very far away (such that their balls of radius r are separate). These two vectors both behave in this way and don't interact, so we can subtract one from the other to get their mean to be 0 — take two test vectors x and y in this way (on far-apart balls) and choose $z = x - cy$ (with $c > 0$) such that $z \perp 1$.

(The diameter of our graph goes to ∞, so by taking r to be slowly growing we can get two such test vectors.) □

Remark 12.17. This is why we didn't take the all-1's vector in the test vector construction — we needed a vector that was quite local.

Remark 12.18. What would happen if you tried to subtract a constant multiple of $\mathbf{1}$ instead? It's true that starting with a vector you can project it onto the orthogonal complement of $\mathbf{1}$; but you lose control of the inner product.

§12.2.2 Traces and closed walks

We'll now see a proof of the corollary using traces and moments (counting closed walks).

We'll first see a much weaker statement, as a warmup. Let's consider $\operatorname{tr}(A^2)$. On one hand, this is the sum of the entries squared, so

$$\operatorname{tr}(A^2) = 2e(G) = dn.$$

On the other hand, it's also the sum of eigenvalues squared; so we have

$$dn = \sum \lambda_i^2 \leq d^2 + (n-1) \max_{i>1} |\lambda_i|^2 .$$

Rearranging, we get the bound that

$$\max_{i>1} |\lambda_i| \geq \sqrt{\frac{d(n-d)}{n-1}} = \sqrt{d} - o(1)$$

(thinking of d as fixed and $n \to \infty$). So we did almost no work and got *some* lower bound, which is already within a factor of 2 of the real bound.

But in this proof, we just looked at the second power of A; a natural question is, if you take a much higher power of A, can you get more? And the answer is yes — that's what we'll do.

The quantity $\operatorname{tr} A^{2k} = \sum \lambda_i^{2k}$ counts closed walks of length $2k$ in the graph. We'd like to get some bound on this quantity. For an arbitrary graph G, the number of walks may really depend on the graph. But if we're working with d-regular graphs, there's a very nice model — the *infinite d-regular tree* $\mathbb{T}_d$.

Claim 12.19 — The number of closed length-$2k$ walks in G starting at a fixed vertex is at least the number of such walks in $\mathbb{T}_d$ (still starting at a fixed vertex).

The reason why this is the case is — imagine doing these walks. On one hand we have G, which has some loops (e.g. a grid); and on the other hand we have the infinite 4-regular tree. Starting with some vertex, you can imagine that at every vertex you have some direction signs that label the outgoing edges; and on the infinite tree there are also direction signs. And you can follow the same signs to get another walk.

If you have a closed walk in the infinite tree, then in order for it to be closed, it has to go back and forth, retracing its steps, to come back to the original starting point. And that walk will also be a closed walk in G. (The converse isn't necessarily true — you might have some loops in G not present in the tree — but the inequality is true.)

Now let's try to lower-bound the number of such walks in $\mathbb{T}_d$. What do such walks look like? We start at v, then make some directional choices, then maybe come back a few steps, then go out in some other direction, and so on; and eventually we come back to where we start.

At each step, there are $d-1$ possible outgoing choices, and one possible way to come back. We also want to record how many steps are outwards and how many steps are inwards. So each step in $\mathbb{T}_d$ is either out, which we'll denote with a $+$, or in, which we'll denote with $-$. If we just record outs and ins, what does such a graph look like? We'll walk out some steps and in some steps, but we can never have more ins than outs (since if they're equal we'll have returned to the origin, and we can only go out). So the number of sequences of $\pm$ counts (paths that go up and down and never go below the horizontal line) is a familiar object in combinatorics, counted by the Catalan numbers — the number of such sequences is

$$C_k = \frac{1}{k+1}\binom{2k}{k}.$$

So we have some count on the number of in-and-out sequences.

And for each out-step, there are either $d-1$ or d choices, depending on whether we're at the origin or not the origin; certainly there's at least $d-1$ choices. So the number of closed walks of length $2k$ in $\mathbb{T}_d$ is at least $(d-1)^k C_k$.

Putting these together, we have

$$\operatorname{tr} A^{2k} \geq n(d-1)^k C_k = \frac{n}{k+1}\binom{2k}{k}(d-1)^k.$$

This gives a lower bound; and then we can continue the calculation we did at the start to get the upper bound

$$\operatorname{tr} A^{2k} \leq d^{2k} + (n-1)\max_{i>1}|\lambda_i|^{2k}.$$

Comparing the two sides, we find that

$$\max_{i>1}|\lambda_i|^{2k} \geq \frac{1}{k+1}\binom{2k}{k}(d-1)^k - \frac{d^{2k}}{n-1}.$$

We were trying to find a lower bound on the left-hand side; and this is some lower bound. The first term is

$$\frac{1}{k+1}\binom{2k}{k}(d-1)^k = (2-o(1))^{2k}$$

as $k \to \infty$. We'll take k to increase with n, but we don't want it to increase too quickly — then the second term (supposed to be the error term) would start dominating. But if we let $k \to \infty$ slowly — e.g. $k = \log\log n$ — then comparing the two sides gives the desired inequality.

Question 12.20. What is the number $2\sqrt{d-1}$?

(It came up in the two proofs, though in different ways.)

It turns out that the answer is that $2\sqrt{d-1}$ is the spectral radius of $\mathbb{T}_d$. This is some infinite graph, so we haven't even defined what a spectral radius means (and we'll be vague); but think of it as a graph, except where we don't get to use the all-1's vector (we're only allowed to use vectors with finite length). Then to find the operator norm we have to go to the next vector.

This fact comes up in both proofs in different ways — in the first proof, we exhibited test vectors which are essentially truncated versions of $\mathbb{T}_d$. And in the second, we essentially computed moments of $\mathbb{T}_d$ by counting walks (which is another way to access its spectral radius). So $2\sqrt{d-1}$ isn't just a random number that comes out of the calculations; it's quite important.

§12.3 Tightness of the bound

Question 12.21. Is $2\sqrt{d-1}$ optimal?

The answer is yes.

Theorem 12.22 (Friedman)
Fix d and $\lambda > 2\sqrt{d-1}$. Then

$$\mathbb{P}(G_{n,d} \text{ is an } (n,d,\lambda)\text{-graph}) \to 1 \text{ as } n \to \infty.$$

Here $G_{n,d}$ is a random d-regular n-vertex graph (we look at the universe of all such graphs and pick one uniformly). So $2\sqrt{d-1}$ is optimal.

This theorem is really hard, and the proof is over 100 pages.

§12.4 Ramanujan graphs

There is another question, which is a very famous open problem:

Question 12.23. Can we get rid of the $o(1)$ — what if we want exactly $2\sqrt{d-1}$?

This leads to the concept of Ramanujan graphs.

Definition 12.24. A *Ramanujan graph* is an (n,d,λ)-graph with $\lambda = 2\sqrt{d-1}$.

So this means all but the top eigenvalue are at most $2\sqrt{d-1}$ in absolute value, without any allowed error term.

Question 12.25. Do Ramanujan graphs exist?

There is a trivial but silly answer — the eigenvalues of K_{d+1} are d, and then -1 repeated d times. But this is silly; the point of the question is, if we fix d and let n grow, can we find infinitely many?

Conjecture 12.26 — For every $d \geq 3$, there exist infinitely many d-regular Ramanujan graphs.

This is a major open problem; there are some known results, and all of them are quite significant.

It's known that the answer is yes if $d-1$ is a prime power — this is a combination of several results. Initially it was proved that these exist when $d-1$ is an odd prime; they also coined the name Ramanujan graphs (not because Ramanujan studied these things, but because their proof constructs such graphs using an explicit Cayley graph, and the proof that it works uses deep number theory related to Ramanujan). This construction was also independently discovered, and was later extended to prime powers.

And this is the entire set of numbers for which we know this conjecture — in particular, the smallest open case is $d=7$.

It is believed that a random d-regular graph will be Ramanujan with some positive probability (empirically it's observed that some fraction are not Ramanujan — they're slightly above the threshold — and some are).

Conjecture 12.27 — For fixed d, there exists $c_d > 0$ such that

$$\mathbb{P}(G_{n,d} \text{ is Ramanujan}) \geq c_d$$

for all sufficiently large n.

But we're very far from being able to show this.

Another very important and recent result shows that *bipartite* Ramanujan graphs exist for every degree.

Theorem 12.28

For every $d \geq 3$, there exist infinitely many d-regular bipartite Ramanujan graphs.

For any bipartite graph, its spectrum is symmetric; if λ is an eigenvalue, so is $-\lambda$. In particular $-d$ is also an eigenvalue; so we exclude it from the definition. So here the definition is just that $\lambda_2(G) \leq 2\sqrt{d-1}$.

This uses a very different argument — it's extremely clever and short. They used a technique called interlacing polynomials that gives a probabilistic construction. It's not a naive random graph — you have to do clever things to construct the graph — but it's not an explicit construction.

Remark 12.29. The authors also thought about, can you use this technique to construct normal Ramanujan graphs? The technique works to handle one eigenvalue, but it can't handle the lowest eigenvalue; so it falls short of being able to prove the conjecture for normal Ramanujan graphs.

§13 October 23, 2023

This Wednesday we have a guest lecture by Fan Chung; she will talk about some Erdős stories (she was a friend of Erdős). Afterwards she will give a seminar talk on quasirandom boolean functions.

On Friday, office hours will be in 2-139 instead of the common room.

§13.1 Graph limits

We'll start a new chapter on graph limits. Today we'll discuss what are graph limits, and the goals and objectives. This is a fairly modern development; its theory was developed by Lovász and coauthors starting in 2003. They were motivated both by pure math considerations and applications; we'll talk about some stories behind them and motivate the subject.

Graph limits give an analytic framework for analyzing large graphs. We saw the regularity lemma and quasirandomness, which are both important ideas we'll use to build up to graph limits. The theme is to have a nice framework to be able to talk about really large objects that have both structure and randomness.

§13.2 Motivation

Suppose we live in a world where we only knew about $\mathbb{Q}$, but not $\mathbb{R}$. (The first chapter of Rudin's analysis textbook starts this way — it builds $\mathbb{R}$ from $\mathbb{Q}$.) Now suppose that someone asks you to solve the following optimization problem.

> **Question 13.1.** Minimize $x^3 - x$ subject to the constraint that $0 \le x \le 1$.

We know how to do this because we know calculus and $\mathbb{R}$, and we (who do know $\mathbb{R}$) know the answer is $x = 1/\sqrt{3}$. But this answer doesn't exist in $\mathbb{Q}$.

So in order to state what the answer is to this question (without knowing $\mathbb{R}$), a reasonable way to do it is to state a *sequence* of rational numbers that converges to this number — and that's basically the best you can do. This is actually one way to *define* the real numbers, as convergent sequences of rationals.

Here's an analogous question in graph theory — suppose we fix $p \in [0,1]$. We want to minimize

$$\frac{\#\text{closed walks of length 4 in } G}{v(G)^4}$$

among all graphs G with at least $pv(G)^2/2$ edges.

We saw this quantity last lecture — it's basically the C_4-count. And in the last chapter, we showed this quantity is at least p^4 (using a Cauchy–Schwarz calculation). This number is best possible, because for a quasirandom sequence it's $p^4 + o(1)$. So the answer to this question is p^4, but there isn't a single graph that attains the answer — this is analogous to the optimization problem in $\mathbb{Q}$.

So what this tells us is that the space of graphs is not complete — you can have sequences that don't have limits in the space of graphs, so to state this answer we need to state a *sequence* of graphs. Wouldn't it be nice to have a single object, analogous to the irrational number $1/\sqrt{3}$, that describes the answer — i.e., the limit of the sequence of quasirandom graphs?

That's one motivation for graph limits — a single object that describes a limit of a sequence of convergent graphs, so we can say that this is minimized at a particular single object.

§13.3 Guiding questions

We'll now state some questions we'll consider in this chapter.

> **Question 13.2.** What does it mean for a sequence of graphs to converge?

We'll in fact have several notions of convergence.

> **Question 13.3.** Are these different notions of convergence equivalent?

> **Question 13.4.** Does every convergent sequence have a limit? If so, what is the limit — what's the object that represents the limit of a convergent sequence?

We'll see the notion of a *graphon*, which we'll define shortly. The word 'graphon' comes from combining the words 'graph' and 'function.'

To extend the analogy in $\mathbb{R}$, in an introductory course in analysis or topolgy we learn about *Cauchy sequences*.

Definition 13.5. A *Cauchy sequence* in a metric space (X, d) is a sequence $x_1, x_2, \ldots$ such that for every ε, there exists N such that $d(x_n, x_m) < \varepsilon$ for all $n, m > N$.

So we can get the elements of the sequence to be arbitrarily close to each other after a sufficiently far point.

Definition 13.6. A metric space is *complete* if every Cauchy sequence has a limit.

For example, $\mathbb{Q}$ is not complete — it has sequences that don't have limits. The *completion* of a space is the smallest complete space that contains the original (there's several ways to define this, but we won't here) — for example, the completion of $\mathbb{Q}$ is $\mathbb{R}$. One main message of this chapter is that the space of graphons will be the completion of the space of graphs. Furthermore, the space of graphons will also be compact.

The first point is analogous to the fact that $\mathbb{R}$ completes $\mathbb{Q}$ (intuitively, it fills in all the holes in the rational numbers). We'll make this statement precise after some time, by developing the right definitions and motivations.

§13.4 Graphons

Graphons will be the central objects of this chapter.

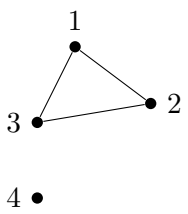
Definition 13.7. A *graphon* is a symmetric measurable function $W\colon [0,1]^2 \to [0,1]$.

Here *symmetric* means that $W(x, y) = W(y, x)$ for all $x, y \in [0, 1]$. We're going to talk a lot about analytic concepts, so it's important to talk about measurability (since we'll be taking integrals). But we won't take this too seriously; we mostly view this as a technicality. There will be times when we talk about Lebesgue measures, but for the most part we can think about it in terms of lengths and areas.

It'll be sometimes useful to consider the domain as some other finite square instead of $[0, 1]^2$, but the math is the same.

There are different opinions about the use of measure theory. There are two camps. One is that it's a very important concept integral to many definitions; if you talk to serious probabilists they will say this, since you can't access some objects in probability theory without the measure theory. But for others, measure theory is mostly a technicality; this is the camp Prof. Zhao is in, so he'll sometimes be cavalier with the measure theory. Some analysts feel that the only time they think about measure theory are when they teach it.

What's the relationship between a graphon and a graph? Here's an easy way to turn a graph into a graphon. Suppose we have a graph:



Then we can write down the adjacency matrix of this graph, which will be

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

We can view this as a pixellated picture where we divide the unit square into a 4×4 grid, and put in a shaded square for every 1 and an unshaded square for every 0.

And now we can view this as a function $W_G \colon [0,1]^2 \to [0,1]$, in this case taking values that are 0 or 1. We call this the *associated graphon* of a graph G.

We don't really care what values we assign at the edges of these squares (it could be either 0 or 1), since we only care about things up to measure 0.

§13.5 Examples

Example 13.8

The *half graph* is the bipartite graph where we only draw the edges from the left to the right whose slope is horizontal or downwards.

The associated graphon contains a staircase in the upper-right and bottom-left corner. As the number of vertices goes to ∞, these steps smooth out into a line segment. So in a certain sense, which we haven't defined yet, this graph turns into the following graphon (and we will see that this graphon is indeed the limit object):

Example 13.9

Consider quasirandom graphs at density $p \in [0,1]$. We want their limit to be the constant graphon that takes value p.

Another example is the *stochastic block model*, which is an important probability model for graph theory. Here we have some red vertices and blue vertices, and we put in random edges; the probabilities of these edges depend on where the endpoints lie. So we may have probability p_r for edges between two red vertices, and p_b for edges between two blue vertices; and p_{rb} for the red-blue edge probabilities. Imagine that these are constants, but the number of vertices grows larger and larger. Then this converges to the graphon which is a step function with four blocks, whose values are given by these probabilities.

This is a generalization of the quasirandom example, where instead of having one density we have several different densities, depending on the types of vertices.

The last example is an important one, since it shows there's a caveat. In a way, the message here seems to be that you take some pixellated picture, and then zoom out (blurring your vision) to see what comes out in the limit. But what about checkerboard graphons (where we consider a sequence as the number of cells increase)? You might say that as we zoom out, this looks like a pretty uniform picture at density $\frac{1}{2}$; so maybe the constant graph at $\frac{1}{2}$ would be an appropriate limit.

But what is this graph? As a graph, the 4×4 checkerboard is $K_{2,2}$, and if we draw a larger checkerboard, then we get the bipartite graph $K_{n,n}$ (where the left and right vertices correspond to even and odd rows and columns).

So by reordering the rows and columns, we see that a different way to draw the same picture is via a two-block graphon. And this picture doesn't even depend on n — we have the same graphon throughout the sequence, and that should be the limit.

So the message of this example is — graphs come with labelled vertices, and for a sensible definition we need to allow relabeling the vertices. That's an important but subtle caveat that one might mistake if we just see the first three examples.

§13.6 Graphon similarity

With these examples in mind, if we're given a graphon (which is some infinite-dimensional object, living in infinite-dimensional space), we'd like to understand some notion of similarity between two such graphons. We'll explain two different approaches to measuring similarity. The first approach is via cut distance; we'll look at this today, and it's a notion motivated by the notions of discrepancy and ε-regularity that we've been looking at. Another notion, which will also be important, concerns homomorphism densities — where we measure these graphs in terms of subgraph counts. A central theorem we'll see the proof of is that these two notions are equivalent — convergence in cut distance and in homomorphism densities are equivalent concepts.

§13.7 Cut distance

Now we'll start to define things more rigorously, starting with cut distance.

Stepping back, suppose we're given two graphs, and we want to say how similar they are to each other. What might be some natural notions of similarity between graphs? For different applications, you may care about different notions. A pretty natural one to start with is the notion of *edit distance* — how many individual changes do we need to go from one graph to another? This is pretty important for certain applications, but it turns out to not be the one relevant for us here. The reason is that for example, we'd like to say two instances of $G(n, \frac{1}{2})$ are very similar to each other, since we want them to converge to the same object. But the edit distance is roughly $\frac{1}{4}n^2$ (half the edges need to be changed to go from one to the other). This is a pretty large distance for two things we want close to each other.

Instead we'll look at a notion of distance motivated by discrepancy.

We'd like G being ε-close to the constant p graphon to mean that

$$|e_G(X,Y) - p\,|X|\,|Y|| \le \varepsilon\,|V(G)|^2$$

for all $X, Y \subseteq V(G)$ — this is the notion we saw in quasirandomness and ε-regularity, and we'd like to use it again to define the notion of cut distance.

This just compares a single graph to a constant graphon. The next thing we want to do is be able to compare two graphs against each other. We'll need to do this in several steps.

For now, we'll look at two graphs sharing the same vertex set. (Vertices are labelled, so we won't have to worry about different relabellings yet.)

Definition 13.10. We say two graphs G and G' on the same vertex set V are *ε-close in cut norm* if the following is true: for all $X, Y \subseteq V$, we have

$$|e_G(X,Y) - e_{G'}(X,Y)| \le \varepsilon\,|V|^2\,.$$

This definition now has the nice property that if we draw two instances of $G(n, \frac{1}{2})$, then they're close to each other in this sense; more generally two instances of $G(n,p)$ are $o(1)$-close with high probability. So we do have the notion of closeness for random graphs that we wanted.

Now we'll define an analogous notion for graphons.

Definition 13.11. The *cut norm* of a measurable function $W\colon [0,1]^2 \to \mathbb{R}$ is defined as

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} W\right|.$$

Here we take sup over all measurable $S, T \subseteq [0,1]$.

(We may drop the word 'measurable' and simply assume it whenever necessary.)

We can see some semantic similarities between the definition for graphs and graphons — given W_G and $W_{G'}$ for graphs G and G' on the same vertex set, we can overlay them and consider $\|W_G - W_{G'}\|$. We're looking at the worst subsets X and Y on the left and the worst subsets S and T on the right; so the two definitions are similar.

Remark 13.12. It's not direct from this definition that G and G' being ε-close is equivalent to $\|W_G - W_{G'}\|_\square \le \varepsilon$ — subsets of $[0,1]$ could pick up fractions of vertices. In this example, it turns out that to maximize the integral you should only look at all or nothing of a vertex (so it's never optimal to take fractional vertices); but that requires a proof.

This definition so far is only defined for vertex-labelled graphs — if we're given *a priori* a labelling on the vertices — and it only allows us to compare two graphs on the exact same set of vertices. But when we want to talk about graphs, we want to talk about unlabelled graphs that allow permuting or relabelling the vertices. To do so, we need to come up with some notion for graphons that also lets us relabel the 'vertices.'

So there's two challenges. First, given two unlabelled graphs, we'd like to find the 'optimal' overlap. Second, what if the two graphs don't have the same number of vertices? We should be able to say a random graph on n vertices is very similar to a random graph on $n+1$ vertices, and so far we don't know how to do this.

The solution is to, instead of comparing them as graphs, first convert them to graphons. Similarly, as graphs it may not seem obvious why $K_{2,2}$ and $K_{4,4}$ are similar, but when we convert to graphons they become the same function.

Definition 13.13. We say that a map $\varphi\colon [0,1] \to [0,1]$ is a *measure-preserving map* if

$$\lambda(A) = \lambda(\varphi^{-1}(A))$$

for all measurable $A \subseteq [0,1]$, where λ denotes the Lebesgue measure.

You should think of the Lebesgue measure in terms of lengths and areas.

This is actually a slightly tricky definition; if you've never seen this before and you try to write down one yourself, you'll write down the wrong thing.

Example 13.14

Consider $\varphi(x) = x + \alpha \mod 1$ (thinking of $[0,1]$ as a circle, or $[0,1]^2$ as a torus); this simply rotates the torus, so it doesn't affect lengths.

Example 13.15

The map $x \mapsto 2x \pmod 1$ is measure-preserving — even though it 'looks' like it dilates by a factor of 2. This is because if we look at some interval A, its inverse has not one but *two* intervals, whose measures add up to that of A.

If we naively wrote down that $\lambda(A) = \lambda(\varphi(A))$ instead, that would *not* include this example; but we do want to include it.

Definition 13.16. We say φ is *invertible* if there exists another measure-preserving map $\psi\colon [0,1] \to [0,1]$ such that $\varphi \circ \psi$ and $\psi \circ \varphi$ are both identity maps outside a set of measure 0.

This definition is sometimes where measure-theoretic technicalities get in the way; you should think of the maps we use as 'bijections' since in your mind you're permuting vertices. (An invertible measure-preserving map is a measure-theoretic analog of a bijection on vertices.)

With that, we can now define the cut metric.

Definition 13.17. Given a map $W\colon [0,1]^2 \to \mathbb{R}$ and a measure-preserving $\varphi\colon [0,1] \to [0,1]$, we write W^φ to denote

$$W^\varphi(x,y) = W(\varphi(x), \varphi(y)).$$

If we think of W as the associated graph and φ as coming from permuting the vertices, then this new graphon is the one associated with the relabelled graph.

Definition 13.18. Given two symmetric measurable functions $U, W\colon [0,1]^2 \to [0,1]$, we define their *cut distance* (or *cut metric*) to be

$$\delta_\square(U,W) = \inf_\varphi \|U - W^\varphi\|_\square,$$

where we take inf over all invertible measure-preserving maps $\varphi\colon [0,1] \to [0,1]$.

Here $\|U - W^\varphi\|_\square$ denotes the cut norm (as defined earlier).

Given graphs G and G' (which are unlabelled graphs, and could even have different numbers of vertices), we define their cut distance to be the cut distance between the associated graphons.

Definition 13.19. For graphs G and G', we define $\delta_\square(G,G') = \delta_\square(W_G, W_{G'})$.

Unlike the previous definition of close in cut norm, this definition works for graphs on different numbers of vertices. Similarly, we can define what it means for a graph to be close to a graphon.

Definition 13.20. For G a graph and U a graphon, we define $\delta_\square(G,U) = \delta_\square(W_G, U)$.

Remark 13.21. Given two graphons, how easy is it to compare the distance between them? If we care only up to a constant resolution ε, we can use the regularity lemma to first regularize up to resolution ε; this allows us to compare at most a constant number of blocks. So for each ε we get a constant; the dependence is not as bad as what we saw earlier in the regularity lemma, but it's still exponential.

Remark 13.22. There is now a subtle but highly nontrivial point — given two graphs on the same number of vertices, what is the cut distance between them? You can imagine two definitions. In one, you first convert to graphons. In another, you consider bijections of vertices, find the optimal overlay, and apply the cut norm. These sound very similar, but they're not exactly the same. With graphons, you might allow fractional matchings — you might map half of the first vertex to one vertex, and half of the first vertex to another. So here instead of bijections between vertices, we're considering fractional perfect matchings.

It turns out this could make a difference — given G and G', the cut distance under this definition could be different from the cut distance if we only took bijections. But there are bounds, which are polynomially far apart — so they're not that different, but they could be different.

§13.8 Convergence

So far, we've defined a notion of cut distance, which we can apply to two graphs, two graphons, or one graph against a graphon. This allows us to define a notion of convergence.

Definition 13.23. A sequence of graphs or graphons *converges in cut metric* if they form a Cauchy sequence with respect to the cut distance $\delta_\square$.

(As before, a Cauchy sequence means that for every ε, if we look far enough out, everything is within an ε-ball of each other.)

Definition 13.24. We say that W_n *converges to* W *in cut metric* if $\delta_\square(W_n, W) \to 0$ as $n \to \infty$.

The first part is about the definition of a *sequence* being convergent — you can talk about a sequence in $\mathbb{Q}$ being convergent without ever talking about $\mathbb{R}$. The second part is about their *limit* — what does it mean for W to be the limit of the sequence?

§13.9 The space of graphons

Now let's talk about the space of graphons (we can put all graphons together and think of them as a space). First, we want to identify graphons with distance 0 (and treat them as a single graphon). This could identify some graphs as well — we saw that the balanced complete bipartite graphs all have the same graphon, and they're treated as a single point in the space of graphons. Modifying a graphon by a measure-0 set also doesn't change the graphon.

Notation 13.25. We use $\widetilde{W_0}$ to denote the set of graphons $W\colon [0,1]^2 \to [0,1]$, equipped with the cut metric $\delta_\square$.

Remark 13.26. The subscript 0 is because $\widetilde{W}$ is sometimes used to refer to *real-valued* graphons (i.e., $[0,1]^2 \to \mathbb{R}$).

Theorem 13.27 (Lovász, Szegedy)
The space $(\widetilde{W_0}, \delta_\square)$ is compact.

How should we think about this? The definition of compactness is that every open cover has a finite subcover. But for Euclidean spaces, being closed and bounded is equivalent to being compact. In a metric space, being compact is also equivalent to being sequentially compact — every sequence has a convergent subsequence with a limit. One way to think about it is that it's closed in some sense — limits exist — and it's 'not too large' (we can't have a sequence that runs away without a limit).

And that's what this theorem is saying — the world of graphs is nice in that graphons fill in all the holes, and also the space of graphs isn't too large. This is counterintuitive — graphs are very versatile and encode a lot of information, but we're saying the space of graphs is not too large. One way to think about this is that every big graph somehow has a small approximation. And we've seen this before in the regularity lemma — the regularity lemma has precisely this message, that if we're willing to tolerate an ε-error then we can approximate every graph by something using only a bounded amount of information. That's a message about the smallness of the space of graphs, which is about compactness.

And in fact, we will prove this theorem using a version of the regularity lemma.

This theorem is very succinct, but the word 'compact' encodes lots of information, similar to that of the regularity lemma.

Remark 13.28. When Prof. Zhao was a freshman, his analysis professor said that in baby Rudin, there's lots of theorems with very short proofs. These proofs are short not because the theorems are easy, but because the definitions are realy good. And that rung true — the definition of compactness is a beautiful definition in that you can define it clearly, but it encodes so much information. And this is an example of that phenomenon.

Remark 13.29. This theory was developed in the 2000s. When Prof. Zhao was a grad student, Lovász said that compactness is so important in analysis that analysts thought they knew all the relevant compact spaces. But this example, coming from graph theory, is a genuinely new type of compact space not analogous to other things we've seen before.

We'll close by proving something which is not this theorem, but is much easier.

Theorem 13.30

The set of graphs is dense in $(\widetilde{W_0}, \delta_\square)$.

Think of graphs as $\mathbb{Q}$ and graphons as $\mathbb{R}$; this says that given a graphon, we can find graphs that approximate the graphon arbitrarily well (we don't have a graphon that's far out there and removed from all the graphs).

Proof. Suppose we start with some arbitrary graphon $W\colon [0,1]^2 \to [0,1]$; we can think of this as having some shaded picture. We'd like to find a graph that approximates this picture to an arbitrarily small resolution. (We're not going to use the regularity lemma.)

We can start by trying to cut our graphon up into grids. But what if our starting graphon is shaded in a diagonal gradient, where it's constant on each antidiagonal and becomes lighter and lighter as it goes down? Dividing this into cells somehow doesn't help that much, since these diagonals still remain diagonals inside the cells. We could do some averaging inside the cells; one issue with that is that maybe inside each cell we have some crazy stuff happening. For example, it could be that inside each cell we have a graph consisting of a small corner; then averaging misses out on the potential for some global effect taking place.

We can try looking at the level sets — we have a full gradient of level sets and we want a discrete approximation, but that's okay, since we can discretize the level sets. So we first convert W into W_1 which only takes values that are integer multiples of ε (so that it only takes $\frac{1}{\varepsilon}$ possible values). Then we have $\|W - W_1\|_1 \le \varepsilon$ (since we lose no more than ε even at every point).

Now our picture looks like a finite number of blobs (possibly with crazy shapes), each of which has some value inside it.

Now each level set is a measurable set, and somewhere in the definition of measurability, it implies that we can approximate the set by grid boxes. We can do this for each set separately, and then obtain W_2 where $\|W_1 - W_2\| \le \varepsilon$ and W_2 is constant on grid boxes.

We don't yet have a graph — right now W_2 is a graphon that looks like a bunch of tiny cells in a grid, and inside each grid is some number.

So we now replace each constant block by a block corresponding to a random (or quasirandom) graph. Our approximations in the first two steps were L^1-approximations (being close in L^1 distance also implies being close in cut norm), and this approximation is a cut norm one.

This shows that for every W and ε, there exists a graph such that $\|W - W_G\|_\square < \varepsilon$. This proves our theorem. □

Question 13.31. How many vertices did we need in our graph approximation G, and does this depend on W?

In this proof, $v(G)$ depends on both W and ε. The step that depends on W is the second one, where we approximated measurable sets using grid boxes. Different measurable sets may require different numbers of boxes. And this was necessary — for any number of boxes, we can design a pretty crazy set that requires that number of boxes to approximate.

But it turns out that you can prove a version of this theorem where the number of vertices in G does *not* depend on the starting graphon, and only depends on W. In fact, it'll turn out that we only need $2^{\varepsilon^{-O(1)}}$ vertices, and this quantity doesn't depend on W.

The fact that you can do this is already encapsulated in the compactness statement. Suppose that we already knew this statement; then we can cover the entire space by a fixed set of ε-balls centered at our graphs (which is a finite set), and the number of balls we need is some bound based on the number of vertices.

We'll prove this in the next couple of lectures, when we develop the tools to prove compactness.

§14 October 30, 2023

We're going to continue our discussion of graph limits. Last Monday we defined a graphon, and the notion of cut norm and cut distance; this defines the space of graphons, under the cut distance. We stated some theorems; most importantly, this space is compact. We also laid out a high level overview of what we'll see in this chapter. We'll look at convergence of graph sequences under two notions. One is under the cut distance — last time we saw how to define the cut distance between two graphs or graphons. Today we'll see a different way to define convergence, through homomorphism densities.

§14.1 Homomorphism density

Definition 14.1. A *graph homomorphism* from F to G is a map $\varphi: V(F) \to V(G)$ such that if $uv \in E(F)$, then $\varphi(u)\varphi(v) \in E(G)$.

In other words, φ is a graph homomorphism if it carries edges to edges. This is a very powerful concept; we'll discuss this for quite a bit.

Notation 14.2. We use $\mathrm{Hom}(F, G)$ to denote the set of homomorphisms $F \to G$, and $\hom(F, G) = |\mathrm{Hom}(F, G)|$ to denote the number of homomorphisms.

Definition 14.3. We define the *homomorphism density* of F in G, or the *F-density* in G, as

$$t(F, G) = \frac{\hom(F, G)}{v(G)^{v(F)}}.$$

The homomorphism density is also the probability that a uniformly random map $\varphi: V(F) \to V(G)$ (allowing repetitions) is a homomorphism $F \to G$.

Example 14.4

We have $\hom(K_1, G) = v(G)$ — we have $v(G)$ ways to map a single vertex to the vertex set of G.

Example 14.5
We have $\hom(K_2, G) = 2e(G)$.

Example 14.6
We have $\hom(K_3, G) = 6\#\text{triangles in } G$.

Example 14.7
We have $\hom(G, K_3) = \#\text{proper 3-colorings of } G$ with red, blue, and green. (Think of red, blue, and green as the vertices of K_3, and the map as an assignment of colors; the constraint of sending edges to edges means no two adjacent vertices should be mapped to red, and so on.)

So homomorphisms encode a lot of interesting statistics on graphs. For today, we'll be mostly interested in the first three cases, where we map a small graph into a larger graph; but it's also interesting to think about mapping a larger graph to a smaller one, which we'll discuss next week.

Example 14.8
We have $\hom(C_4, G) = \#\text{length-4 closed walks in } G$.

In particular, note that this is not the number of 4-cycles, since in a homomorphism we can reuse vertices — it's not necessarily injective. But when we think of G as having a lot of vertices, the number of non-injective homomorphisms will have smaller order; so for large G (meaning that $v(G) \to \infty$), we have $t(F, G) \approx t^{\text{inj}}(F, G)$ (where we only count subgraphs instead of homomorphisms).

So when we think about large G, we can think about this as the number of 4-cycles (provided that we're considering dense graphs, with quadratically many edges).

§14.2 Homomorphism density for graphons

Recall that graphons are measurable symmetric functions $W\colon [0,1]^2 \to [0,1]$, where *symmetric* means that $W(x, y) = W(y, x)$. Sometimes it will be useful to allow W to take real values; but that's just a small difference.

We'll first define F-density in a graphon for the specific case $F = K_3$: we define

$$t(K_3, W) = \int_{[0,1]^3} W(x, y)W(x, z)W(y, z)\, dx\, dy\, dz.$$

We have three variables x, y, and z, which we think of as vertices; these so-called vertices are elements of $[0, 1]$. Then a triangle roughly corresponds to $W(x, y)W(x, z)W(y, z)$. So that's the triangle density in a graphon.

Last time, we defined the notion of an associated graphon — given a graph G, we can turn it into a graphon W_G. (We look at the adjacency matrix, and view it as a black-and-white picture on top of the unit square.) This definition is consistent with the associated graphon transformation — we have $t(K_3, G) = t(K_3, W_G)$.

We can extend this formula from triangles to general graphs.

Definition 14.9. Let F be a graph, and W a graphon (or $\mathbb{R}$-valued). Then

$$t(F, W) = \int_{[0,1]^{V(F)}} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i \in V(F)} dx_i.$$

This is a generalization of the formula we just saw for triangle density — we have variables x_i for each vertex of F, and we multiply W over all edges. The same formula that $t(F,G) = t(F,W_G)$ remains true — if we start with a graph and convert it to a graphon, the notions of F-density for graphs and graphons are the same.

§14.3 Convergence

Now we can define a notion of convergence, which we'll call *left convergence*.

Definition 14.10. We say a sequence of graphons W_n is *left convergent* if for every graph F, the numbers $t(F,W_n)$ converge as $n \to \infty$. We say this sequence left-converges to W if for all F we have $\lim_{n\to\infty} t(F,W_n) = t(F,W)$.

So this tells us what it means for a sequence of graphons (or graphs) to converge — for *every* F, we look at if the F-densities converge. (It's called *left* convergence because it's about homomorphism densities from the left.)

We can give the same definition for sequences G_n of graphs.

This gives two definitions — the first is for what it means for a sequence to converge. In an abstract topological or metric space, it's possible to have a sequence that converges without a limit (for example, in $\mathbb{Q}$ you can have a sequence converging to $\sqrt{2}$, without $\sqrt{2}$ being in the space); and the second is about a sequence converging *to a limit*.

One of the big theorems we'll see in this chapter is the equivalence of convergence. We've seen two notions of convergence — one based on cut distance, and one based on homomorphism densities. The following theorem says that these two notions are equivalent.

Theorem 14.11

Left convergence is equivalent to being a Cauchy sequence with respect to $\delta_\square$.

There is a subtletly here — this theorem has two parts. If we have a sequence, then it left-converges if and only if it's a Cauchy sequence with respect to $\delta_\square$. The second part is that if we have a *limit*, left-convergence to that limit is equivalent to convergence in $\delta_\square$ to that limit.

Remark 14.12. There's also a notion of right-convergence, which is roughly what happens if we test homomorphism densities from the right (but is more subtle than that).

One direction is easy (convergence in $\delta_\square$ to left-convergence); the other is much harder.

So far, this theorem doesn't tell us about the existence of limits — given a convergent sequence, we don't know that there is a graphon representing the limit. But we can show this as well.

Theorem 14.13

Every left-convergent sequence of graphs or graphons left-converges to some graphon.

So if we have a convergent sequence, then there exists a limit object. This theorem was originally proved without the equivalence; but we'll do it through the path of least resistance. Last time we stated without proof that the space of graphons is compact under the cut metric; part of that definition is that subsequential limits always exist. Closedness encapsulates the existence of limits; so that'll give us existence of limits under $\delta_\square$, which will imply the existence of limits under left-convergence.

§14.4 Counting lemma

Today we'll prove the easier implication in the equivalence. This easier implication amounts to essentially a counting lemma, which we already saw in the chapter on the regularity lemma.

There, the counting lemma took the following form: if we have three vertex sets forming ε-regular pairs, and we're trying to count triangles, the counting lemma says this number of triangles should be about what we expect from if we had random graphs between the parts (we proved a lower bound).

This counting lemma will be similar in spirit.

Theorem 14.14

Let F be a graph, and let $W, U : [0,1]^2 \to [0,1]$ be graphons. Then

$$|t(F,W) - t(F,U)| \le e(F)\delta_\square(W,U).$$

So if U and W are similar in the sense of cut distance, then their F-densities should also be similar.

This immediately gives the left implication — if we have a Cauchy sequence with respect to the cut distance, then $\delta_\square(W_n, W_m)$ gets closer and closer to 0, so $t(F,W_n)$ and $t(F,W_m)$ get closer and closer to each other.

Proof. It's enough to prove that the left-hand side is at most $e(F)\|W-U\|_\square$ (the difference between cut distance and cut norm is that in the cut distance we are allowed to permute variables), since we can replace U by $U^\varphi(x,y) = U(\varphi(x),\varphi(y))$. This transformation doesn't change F-density (permuting vertices doesn't change the triangle density); so we can make this replacement and take $\inf_\varphi$.

To prove this inequality, it'll be helpful to reformulate the cut norm in a way that's a bit more useful. Recall that for a symmetric measurable function $W : [0,1]^2 \to \mathbb{R}$, we defined

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W \right|$$

(where we take sup over all *measurable* $S, T \subseteq [0,1]$). We'll reformulate it as the following.

Lemma 14.15

We have

$$\|W\|_\square = \sup_{u,v} \left| \int W(x,y)u(x)v(y)\,dx\,dy \right|,$$

where the sup is over measurable *functions* $u, v : [0,1] \to [0,1]$.

So the statement of the lemma says that we can replace subsets by functions.

Proof. We can rewrite the original definition as

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int W(x,y)\mathbf{1}_S(x)\mathbf{1}_T(y)\,dx\,dy \right|.$$

So one direction is easy — the functions $\mathbf{1}_S$ and $\mathbf{1}_T$ are possibilities for u and v, so the second is at least the first by definition; the content of the lemma is the reverse inequality.

The point si that the expression inside $|\cdot|$ is bilinear in u and v. (You can think about matrix-vector multiplication instead of an integral; it's essentially the same point.) So if we hold u fixed and ask what v we should choose, then at every point we should look at the rest of the integral, and either take v to have

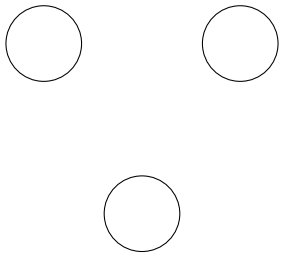
value 0 or 1 (in order to maximize or minimize the function). In other words, for a bilinear function, the maxima are always attained at the extrema — so if we hold u fixed, then the maximum is attained when v is $\{0,1\}$-valued. And the same is true in reverse. So even though the second expression looks more general because it allows u and v to take *any* values in $[0,1]$, in reality the optimum is always attained when they take values in $\{0,1\}$, in which case we get back the original definition. □

Now we are ready to prove the counting lemma. We'll do this just in the case of triangles; the general proof is exactly the same, but with hairier notation. For triangles, the statement is that for graphons W and U, we have

$$|t(K_3, W) - t(K_3, U)| \le 3 \|W - U\|_\square .$$

(This is what we want to prove.) We'll first do this visually: imagine we have three circles with W, W, and W between them, so that initially the triangle density is $\int W(x,y)W(x,z)W(y,z)$. We'll then imagine replacing one of the W's by U's, so we now have $\int U(x,y)W(x,z)W(y,z)$. And we'll show that these two quantities are very close to each other — if $\|W-U\|_\square \le \varepsilon$, we will show that the difference between these two quantities is at most ε.

And then we can keep going — we can replace another W-factor with a U, again losing at most ε. And then we can replace the final W-factor with U as well, also losing at most ε.



So once we prove each of these approximations, that gives the claim. We'll prove one of these; the others will be the same.

The difference between the first two quantities is

$$(\dagger) = \int W(x,y)W(x,z)W(y,z) - \oint U(x,y)W(x,z)W(y,z).$$

We can rewrite this as

$$(\dagger) = \int_{[0,1]^3} (W-U)(x,y)W(x,z)W(y,z)\,dx\,dy\,dz.$$

Imagine fixing z; then we have an integral with $W-U$, some function of x, and some function of y. But that's the definition of the cut norm — so for each *fixed* z, the integral has absolute value at most $\|W-U\|_\square = \varepsilon$. Then we can integrate over z to get that $|(\dagger)| \le \varepsilon$.

That's essentially the entire proof (we do this two more times, and then we're done); we've proven that the difference between triangle densities in U and W is at most 3ε. Of course, for general F we can do the same thing, replacing one edge of F at a time; the factor we lose is $e(F)$. □

This is a counting lemma. The proof, interestingly, is quite different from the one we saw in the regularity chapter; there we embedded one vertex at a time. Here we're doing something which doesn't have as much of a combinatorial interpretation; we're somehow swapping the graph out from W to U, one edge of F at a time. But the analytic proof is very short — this is one of the powers of the analytic approach.

This also gives us a *two-sided* bound. And it also fills in a gap we had in the previous chapter, when proving the Chung–Graham–Wilson theorem on equivalence of quasirandomness (that discrepancy implies subgraph counts). This is implied by this counting lemma — if we take U to be the constant-p graphon, then the content of this claim is precisely that discrepancy implies subgraph counts.

§14.5 Weak regularity

Last Monday, we said that the theorem on compactness of the graphon space is somehow similar to the statement of graph regularity — the space of graphons or graphs is not that large, even though it seems like there's lots of possibilities for what could be in a graph. We'll now explore this more precisely in the context of graphons, and develop a regularity lemma that we'll use to prove this compactness. But this regularity lemma is also interesting on its own.

We'll first define weak regularity for graphs.

Definition 14.16. Given a graph G and a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$, we say $\mathcal{P}$ is *weak ε-regular* if for all $A, B \subseteq V(G)$, we have

$$\left| e(A,B) - \sum_{i,j=1}^{k} d(V_i, V_j) \, |A \cap V_i| \, |B \cap V_j| \right| \le \varepsilon v(G)^2.$$

For all vertex subsets A and B, imagine trying to estimate $e(A,B)$ based on density information just given by the partition. To do so, we can enumerate over all pairs of parts, and look at how many vertices of A lie in V_i and how many vertices of B lie in V_j, and then look at the edge density between them. So we have a graph G chopped up into a few pieces, and we have some A and B; and we want to know the number of edges between them. We estimate the number of edges between one chunk of A and one chunk of B by looking at the densities between the parts in $\mathcal{P}$; and we add these up, and that gives us a naive estimate. The notion of weak regularity is that this naive estimate is never too far off.

We'll take some time to interpret this statement and discuss how it's different from graph regularity as seen earlier. In graph regularity, the condition was that between most pairs of parts we have ε-regularity; that means we can look at subsets of vertices in each part, and say something about the bipartite graphs between them. We have some exceptions — irregular parts where we can't say anything — but other than those, we can say something about vertex sets within parts.

The notion of weak regularity is different — if we have a lot of parts (though a bounded number), if we just look at vertex subsets within parts, these are pretty small — they'll have sizes proportional to $\frac{1}{m}$, which is much smaller than the error estimate. So intuitively, weak regularity is about *global* discrepancies — it's about discrepancies when A and B are *large* (fairly linear-sized). So we'll have a big A which crosses many parts and a big B which crosses many parts, and we'll look at what happens between them. This is a weaker notion — of course if you have finer control of what happens between parts, you can say more. But this also gives better bounds — the graph regularity lemma required a tower of exponentials, but here we'll see we only need a single exponential. This was originally developed for algorithmic applications — if you have much better control on the number of parts, then you get algorithms with reasonable runtime.

Theorem 14.17 (Weak regularity lemma for graphs)
For every ε, every graph has a weak ε-regular partition into at most $4^{1/\varepsilon^2}$ parts.

So similar to graph regularity we have a bounded number of parts; but now this bound is much smaller. In this class we won't be too concerned with the quantitative bounds, but it's nice to know.

And this also leads us to a cleaner notion of how to work with graphons. Right now the definition we have looks very combinatorial, but we can translate it to something more analytic, that's easier to work with for graphons.

To do so, we'll need to introduce some definitions.

Definition 14.18. For a symmetric measurable function $W: [0,1]^2 \to \mathbb{R}$ and a partition $\mathcal{P} = \{S_1, \dots, S_k\}$ of $[0,1]$ (into measurable sets), we define the *stepping operator* $W_{\mathcal{P}}$ as the function whose value in $S_i \times S_j$ is the average of W in this cell.

So we start with W, and then cut up both the x-axis and y-axis according to $\mathcal{P}$; this splits $[0,1]^2$ into a bunch of rectangular cells. And in each cell, we average the value of W inside that cell, and replace W with this average value.

We call this a *step graphon*, since it consists of several steps.

Definition 14.19. We say a partition $\mathcal{P}$ of $[0,1]$ is *weak ε-regular* if $\|W - W_{\mathcal{P}}\|_{\square} \le \varepsilon$.

This notion is consistent to the definition for graphs. We should at least see semantic similarities — we're taking the worst A and B, while the cut norm essentially encodes taking the worst subsets. There's a subtlety with the possibility of fractional vertices, but this doesn't really matter for the same reason as in our reformulation of the cut norm.

Theorem 14.20 (Weak regularity lemma for graphons)
For every $0 < \varepsilon < 1$, every graphon has a weak ε-regular partition into at most $4^{1/\varepsilon^2}$ parts.

Remark 14.21. Does the graphon version of the regularity lemma imply the graph version (stated as is)? The answer is no — a partition of $[0,1]$ might not respect vertices. In the associated graphon, each vertex becomes a segment; but when we take a partition of $[0,1]$, there's no guarantee that a segment hasn't been broken up further. So the partition we end up with when applying the weak regularity lemma for graphons might not be a partition of vertices.

But this doesn't actually matter. For one thing, we can just redo the same proof. More abstractly, in the definition of a graphon we can replace $[0,1]$ with any probability space, for example a finite one; then the same proof works, and that forces you not to break up vertices.

When we proved the graph regularity lemma in Chapter 2, we used energy — the squared mean sum of all the edge densities between parts. We'll see a related quantity here, which is actually even cleaner. In this case, the energy will simply be the L^2 norm of the step graphon (or rather, its square) — we define the *energy* as

$$\|W_{\mathcal{P}}\|_2^2 = \int_{[0,1]^2} W_{\mathcal{P}}(x,y)^2 \, dx \, dy.$$

So we look over all squares, take the mean squared of W, and average appropriately.

Since we're talking about energy, we'll also talk about inner products: we define

$$\langle W, U \rangle = \int WU = \int W(x,y)U(x,y) \, dx \, dy.$$

When we talked about energy in the graph regularity chapter, we proved several lemmas that said that if we refined a partition then energy can never go down, and if the original partition is irregular then it has to go up by quite a bit. We'll have a similar lemma here.

Lemma 14.22 (Energy increment)
Let W be a graphon, and suppose that $\mathcal{P}$ is a finite partition that is not weak ε-regular for W. Then there exists a partition $\mathcal{P}'$ obtained by subdividing each part of $\mathcal{P}$ into at most 4 parts such that $\|W_{\mathcal{P}'}\|_2^2 \ge \|W_{\mathcal{P}}\|_2^2 + \varepsilon^2$.

So we have a definite increase in energy after doing this refinement.

Proof. Because $\mathcal{P}$ is not weak ε-regular, we have $\|W - W_{\mathcal{P}}\|_{\square} > \varepsilon$. By the definition of the cut norm, there exist $S, T \subseteq [0,1]$ such that

$$|\langle W - W_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle| > \varepsilon.$$

(A small caveat is that we can't replace both inequalities with $\geq$ — the cut norm is defined as a supremum, and the supremum might not be attained; whether a supremum is attained is a very delicate issue in the theory of graphons, which we won't get into.)

Let $\mathcal{P}'$ be the refinement of $\mathcal{P}$ obtained by introducing S and T — we throw in these two sets, and then cut up everything further. We know that each part of $\mathcal{P}$ is subdivided into at most 4 parts.

We know that $\langle W_{\mathcal{P}}, W_{\mathcal{P}}\rangle = \langle W_{\mathcal{P}'}, W_{\mathcal{P}}\rangle$, because on every part of $\mathcal{P}$ we are subdividing further, but we're just taking weighted averages — if we look at a part of $\mathcal{P}$ and a part of $\mathcal{P}'$, if we redistribute mass but the average is the same, then the two are equal on every part of $\mathcal{P}$. This means

$$\langle W'_{\mathcal{P}} - W_{\mathcal{P}}, W_{\mathcal{P}}\rangle = 0.$$

And now we're back to the Pythagorean theorem — this implies that

$$\|W_{\mathcal{P}'}\|_2^2 = \|W_{\mathcal{P}}\|_2^2 + \|W_{\mathcal{P}} - W_{\mathcal{P}'}\|_2^2.$$

So we just need to show that the second term is large.

To see this, we have

$$\|W_{\mathcal{P}} - W_{\mathcal{P}'}\|_2 \geq \|\langle W_{\mathcal{P}'} - W_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle\|$$

by Cauchy–Schwarz (think about inner products between vectors). On $S \times T$, we can replace $\langle W_{\mathcal{P}'}, \mathbf{1}_{S\times T}\rangle$ with $\langle W, \mathbf{1}_{S\times T}\rangle$, since $\mathcal{P}'$ is defined by introducing S and T — S and T are each unions of parts of $\mathcal{P}'$, so both represent the same weighted average. So then we can replace this expression with

$$|\langle W - W_{\mathcal{P}}, \mathbf{1}_{S\times T}\rangle|,$$

which we assumed is greater than ε. This is what we wanted, and proves the claim. □

Remark 14.23. This is basically the same idea we saw in the second chapter, but much more condensed (and done analytically).

Remark 14.24. How does the Cauchy–Schwarz inequality work? Cauchy–Schwarz tells us that if $\|u\| = \langle u, u\rangle$ denotes the length of a vector u, then we have $\|u\|\,\|v\| \geq |u, v|$. Here we use the fact that $1 \geq \|\mathbf{1}_{S\times T}\|$.

Now we'll finish the proof of the weak regularity lemma. We'll actually prove something slightly stronger.

Theorem 14.25

Let $0 < \varepsilon < 1$. Let $\mathcal{P}_0$ be a finite measurable partition of $[0,1]$. Then every graphon W has a weak ε-regular partition $\mathcal{P}$ such that $\mathcal{P}$ refines $\mathcal{P}_0$ and each part of $\mathcal{P}_0$ is partitioned into at most $41/\varepsilon^2$ parts under $\mathcal{P}$.

In the version we've stated before we start with the trivial partition, but we can really start with any partition; we'll need this next time.

Remark 14.26. We saw a similar statement when discussing the strong regularity lemma; Jacob Fox called this the *strong weak regularity lemma*.

Proof. We'll obtain this partition by repeatedly applying the energy increment step. Starting with $i = 0$, if $\mathcal{P}_i$ is weak ε-regular then stop. Otherwise, by energy increment there exists a measurable partition $\mathcal{P}_{i+1}$ refining each part of $\mathcal{P}_i$ into at most 4 parts, such that $\|W_{\mathcal{P}_{i+1}}\|_2^2 \geq \|W_{\mathcal{P}_i}\|_2^2 + \varepsilon^2$. (We then increment i by 1 and do the same thing.)

This process must terminate after at most $1/\varepsilon^2$ steps, since $\|\cdot\|_2^2 \in [0,1]$; this gives us the partition we want. □

The overal scheme is very similar to the regularity proof we saw in the second chapter; just the specific goals are different.

Remark 14.27. Is there a version of this if instead of a graphon, we allowed a symmetric measurable function $W\colon [0,1]^2 \to \mathbb{R}$? The proof is essentially limited by the L^2 norm, so we might need to adjust. But there are notions of sparse regularity, which is really about what happens for unbounded graphons. There we take the adjacency matrix and multiply by $\frac{1}{p}$, creating much bigger entries. There does exist a version of sparse graph regularity.

Remark 14.28. The original motivation for weak regularity (due to Frieze–Kannan) was for MaxCut. An important problem in computer science is to compute the size of the maximum cut — a partition of the graph into two parts so that there are as many edges between them as possible. This is a hard problem; there is a famous algorithm giving an approximation of around 0.878. There are some assumptions under which you should not be able to do better; and it's known that you cannot do much better than 94% assuming $\mathsf{P} \neq \mathsf{NP}$.

But for dense graphs, the situation is different — you can get arbitrarily good *additive* approximations using the regularity lemma. (We think of ε as our error; we pick a constant ε, and take a regularity partition. Then we have a finite number of parts, so we can brute force over them, since this is determined by a finite amount of data.)

So for dense graphs there are fast additive approximations for MaxCut.

§15 November 1, 2023

We've been talking about graph limits for the last couple of lectures. We've stated a lot of definitions and some of our main results. The two results we'll prove today are the following.

Theorem 15.1

The space of graphons $(\mathcal{W}_0, \delta_\square)$ is compact.

Theorem 15.2

Left-convergence and convergence with respect to $\delta_\square$ are equivalent.

Today we'll prove these results, and develop some tools to do so.

§15.1 Compactness

First we'll prove compactness. This has a couple of components — compactness is about the space being intuitively not too large (this should coincide with our intuition about the regularity lemma), but it's also about being closed (limits always exist). So we need a way to produce limits — given a sequence of graphs or graphons, we need a way to construct a limit. The tool we'll use to do so is a nice result from probability called the martingale convergence theorem.

§15.1.1 Martingale convergence theorem

Theorem 15.3 (Martingale convergence theorem)

Every bounded discrete-time (real-valued) martingale converges with probability 1.

We'll first recall what a martingale is and then prove this theorem.

Definition 15.4. A discrete-time martingale is a random real sequence X_0, X_1, X_2, ... such that:

- We have $\mathbb{E}[X_{n+1} \mid X_0, X_1, \ldots, X_n] = X_n$.
- We have $\mathbb{E}\,|X_n| < \infty$.

The first point (the more important one) is a bit of shorthand — on the right-hand side we have a random variable. But this means that if we know the first n terms and ask to reveal the $(n+1)$st, then the expectation is the same as the nth term. The second condition is a technical one we won't have to deal with today.

So that's a martingale. We may have seen martingales in some form or the other; they're a pretty basic concept in probability. But here are a few examples that are good to keep in mind as we prove the result and use the theorem.

Example 15.5

Suppose we have a sequence Z_1, Z_2, ... of i.i.d. mean zero random variables, with $\mathbb{E}\,|Z_i| < \infty$. Then the running sum $X_n = Z_1 + Z_2 + \cdots + Z_n$ is a martingale.

Example 15.6

Suppose we're in a fair casino, where you go in with a pot of money and can make real-time bets (at each point you get to decide how much money to bet, which could depend on prior information). (For example, you might start with \$50, and stop once you get to either \$0 or \$60). In such a betting strategy, if X_n is your balance after the nth bet, then X_0, X_1, ... is a martingale. (No matter what's happened before, in a fair casino, no matter what you do the expected change is 0.)

Remark 15.7. The word martingale was originally used to describe a specific kind of betting strategy, where you go to a casino and want to win \$1. You first bet \$1; if you win then you go home. If you lose you bet \$2. If you win you again go home; if you lose then you bet \$4, and so on. The argument is that at some point you'll win, and then you get a \$1 gain and go home.

What's the problem with this strategy? You might run out of money.

A simple but important observation is that $\mathbb{E}X_n = \mathbb{E}X_0$ (typically X_0 is deterministic) — the expectation doesn't change in a martingale as we progress through the steps (there's no free lunch).

Example 15.8

In a *Doob martingale* we have some hidden variable X (which we may never get to see), and X_n is the expectation of X after revealing all information available up to time n.

For example, suppose you want to know the temperature next Monday — this is some random variable. As each day progresses, based on new available information you have a better estimate. And that change in estimates over time shoudl be a martingale.

(We're being somewhat vague here. If you want to define a martingale much more precisely in probability theory terms, you need to talk about filtrations; we won't talk about this.)

In the third example, we should have $X_n \to X$ as $n \to \infty$ with probability 1 — as you reveal all the information available, X_n should converge to the thing it's supposed to capture (i.e., the random variable X). So that's some intuition.

Example 15.9

As another real-world example, every 4 years there's a presidential race in the U.S., and people bet which candidate will win (expressed as some probability). Based on all the available information up to some point, there's some inferred probabilities. In theory, that number should move like a martingale; and it should converge to 0 or 1 by whenever the election is declared.

Now let's prove the martingale convergence theorem. The proof is actually kind of nice — we'll use the betting strategy interpretation of a martingale.

Theorem 15.10 (Martingale convergence theorem)

If $X_0, X_1, \ldots \in [0,1]$ is a martingale, then the sequence converges with probability 1.

The first time you see this result, it's a bit confusing — the first example is not convergent. (If we flip coins and add ± 1 based on the result, that's really not convergent.) But it's also not bounded, so that's not a problem to begin with. A better model is the Doob martingale, where you do end up with something bounded, though you don't know what it is.

Proof. Suppose that there is some instance $X_0, X_1, \ldots \in [0,1]$ that does not converge. Then there exists a pair of *rational* numbers $0 < a < b < 1$ such that the sequence *upcrosses* $[a,b]$ infinitely many times.

To define *upcrossing*, imagine we have a picture with two horizontal lines at a and b. We can view our martingale as a sequence of discrete points. And up-crossing means that the sequence starts below a, and then ends up above b. If you have any real sequence that doesn't converge, then there has to be a pair of horizontal lines that it upcrosses infinitely many times (you can deduce this from basic definitions of convergence). When we dip below a we start tracing a line, and we stop as soon as we're above b. Each of these is an upcrossing, and we must have infinitely many.

Claim 15.11 — For fixed $a < b$, we have $\mathbb{P}(\text{upcross } [a,b] \text{ infinitely many times}) = 0$.

The theorem then follows from the claim, since we can take a union over all $a, b \in \mathbb{Q}$ with $a < b$. There are countably many, and a countable union of zero-probability events also has probability 0. So if we don't converge then this event holds, and this event holds with probability 0.

It remains to prove this claim. To do so, consider a betting strategy — imagine that X is a sequence of stock prices. We're going to basically buy, hold, and sell, according to the following strategy: if X_n dips

below a, then we buy and hold 1 share (but not more). And if X_n reaches above b, then we sell the share (if we have it).

So at each time-step we see the price and decide what to do. When we dip below a we buy and hold this 1 share; we keep on holding it for the duration of the line, and when we get above b then we sell it.

Let's say we start with $Y_0 = 1$ as our budget, and let Y_n be the value of your portfolio (meaning the amount of cash you have, plus the value of a share if you're holding one). Then Y_n is a martingale — this is a fair casino. (It doesn't matter what you do; the expected change in Y in the next time period is 0.) And $Y_n \geq 0$ — you start with \$1 (and the price is always at most \$1), and you always sell for more than what you paid for, so your cash is never negative.

So if we buy and sell k times up to time n, then $Y_n \geq k(b-a)$. (You may have some left over, but you at least have this much in value.) So then

$$\mathbb{P}(\geq k \text{ upcrossings of } [a,b] \text{ up to time } n) \leq \mathbb{P}(Y_n \geq k(b-a)) \leq \frac{\mathbb{E}Y_n}{k(b-a)} = \frac{1}{k(b-a)}$$

(since Y_n is a martingale, so $\mathbb{E}Y_n = Y_0 = 1$). In particular, this decreases as k increases; so

$$\mathbb{P}(\geq k \text{ upcrossings}) = \lim_{n\to\infty} \mathbb{P}(\geq k \text{ upcrossings up to time } n) \leq \frac{1}{k(b-a)}$$

(the first equality is by the monotone convergence theorem). Taking $k \to \infty$ gives the claim. □

§15.1.2 Proof of compactness of space of graphons

Now we're going to use the martingale convergence theorem to prove compactness of the space of graphons. The reason it's relevant is that it provides us with a limit — we know that if a sequence of reals converges then it has a limit, and that's the way we'll construct the limiting graphon.

The space of graphons is a metric space — it comes with a metric, namely the cut distance. And in a metric space, compactness is equivalent to sequential compactness — i.e., that every sequence has a subsequence converging to some limit point.

So let's start with some sequence of graphons W_1, W_2, W_3, ...; we want to show that there exists a limit point with respect to $\delta_\square$.

The first step is that we're going to regularize the sequence by applying an appropriate form of the regularity lemma. By applying the weak regularity lemma, we can obtain a sequence of partitions $\mathcal{P}_{n,k}$ of $[0,1]$ (where everything is measurable) such that the following holds:

(a) $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$.

(b) We have $|\mathcal{P}_{n,k}| = m_k$, where m_k is some quantity only depending on k.

(c) W_n is regularized by $\mathcal{P}_{n,k}$ — letting $W_{n,k} = W_{\mathcal{P}_{n,k}}$, we have $\|W_n - W_{n,k}\|_\square \leq \frac{1}{k}$.

The regularity lemma indeed gives us these properties. In the statement we had an upper bound on the number of parts; but we can allow empty parts, so we can add in empty parts to get an equality.

Essentially, what's happening is that we start with a sequence of graphons, and we try to regularize it one row at a time. We take W_1 and regularize it first to $W_{1,1}$, then futher to $W_{1,2}$, then to $W_{1,3}$, and so on. And we do the same for W_2, W_3, and so on. (At each step, the partition we get is a refinement of the previous one.)

The next step is that we're going to pass to some subsequence. We'll start with a sequence, and our goal is to find a subsequence with a limit point. So we'll frequently pass down to some infinite subsequence, forget about the things we didn't keep, and then relabel everything back to W_1, W_2, ...; we'll keep passing and relabelling without further comment.

Initially, each partition $\mathcal{P}_{n,k}$ partitions $[0,1]$ into some number of parts; these parts are just measurable subsets, not contiguous intervals. But by passing to subsequences, we can ensure the following:

- For all k and $i \in [m_k]$, we have $\lambda(i\text{th part of } \mathcal{P}_{n,k}) \to \alpha_{k,i}$. (There are only countably many such choices for each k and i; we go through these countably many choices, and for each we restrict to some subsequence where this converges — which we can do because it's a real bounded sequence.)
- Similarly, for all k and all $i, j \in [m_k]$, the value of $W_{n,k}$ on the box

$$(i\text{th part of } \mathcal{P}_{n,k}) \times (j\text{th part of } \mathcal{P}_{n,k}) \to \beta_{k,i,j}.$$

  (Imagine writing down the edge density matrices for our partitions; then we have a finite set of real numbers, and we want to make sure they converge.)

So we have W_n, and we regularize it to $W_{n,k}$. And after passing through sequences, what is $W_{n,k}$? We've chopped up the interval into m_k parts, and we want that the lengths of each part converge to some value α_i, and the value inside converges to some number β_{ij}. And this is going to be our model of the limit.

We're going to let U_k denote this graphon — so U_k is the graphon that is our model of what the limit along the subsequences is, after passing to a subsequence where all the lengths converge and all th estep values converge.

Our mental model should be that all of these are step graphons; and after many rounds of passing to subsequences, the vertical sequences converge to U_1, U_2, U_3, ... in $\delta_\square$ (after rearranging intervals). We can do this because $W_{1,1}$, $W_{2,1}$, ... are finite-dimensional objects (regularization causes finitely many parts), and finite-dimensional objects in bounded real space are sequentially compact.

So we've constructed step graphons U_k such that $\delta_\square(W_{n,k}, U_k) \to 0$ as $n \to \infty$ for every k. Now we have $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$ (when we take a finer and finer partition, if we look one step further but then perform the kth stepping operator, we should go back to the kth step). So the same equality should be true for the U's as well — we also have $U_k = (U_{k+1})_{\mathcal{P}_k}$, where $\mathcal{P}_k$ is the partition of $[0,1]$ into parts of lengths $\alpha_{k,i}$. (In the construction of U_k, we partition the lengths using $\mathcal{P}_k$, and insert the limiting values inside each cell.)

Now we have a sequence U_1, U_2, ...; let's think about what's going on with it. We have U_1 consisting of $|\mathcal{P}_1| = 1$; here nothing happens, and it's literally a single constant $\beta_{1,1}$. And then we have U_2 where this cell has been broken into some pieces, where we have some values in these cells that overall have a weighted average equal to the original cell. So we start with some mass, and then we redistribute it into the various cells (satisfying conservation of mass). And then U_3 is a further redistribution of the mass inside each U_2-cell into the corresponding U_3-cells. So at each step, the mass gets distributed in some way; not necessarily evenly, but there is conservation of mass.

An important but somewhat tricky step is that this is a martingale! What do we mean by this? One way to interpret it is to consider the sequence of random variables $U_k(X, Y)$, where (X, Y) is a uniform random point in $[0,1]^2$ (we fix a random point, and then read out the values at that point along the sequence); that sequence is a martingale. A more natural way to view this is that when we talk about random variables, there's an underlying probability space; here that's $[0,1]$. And our random variable is a *function* on a probability space, and that function is U. And we're saying this sequence of functions forms a martingale. Why? If we knew the value of U_2, that tells us what cell we're in. And (X, Y) is uniformly distributed in this cell (as far as we know); we know that by passing to the next step, we might get a different value, but by conservation of mass the expectation doesn't change.

Since U_k is $[0,1]$-valued, by the martingale convergence theorem there exists a graphon U such that $U_k \to U$ pointwise almost everywhere as $k \to \infty$ (this is what it means to converge with probability 1 — it converges pointwise almost everywhere as a function on the unit square). And that's going to be our limit.

Claim 15.12 — The relabelled subsequence W_n converges to U in $\delta_\square$.

(This is not the original sequence we began with — we did a lot of passing to subsequences and relabelling — but after that we claim that our subsequence converges to U.)

Proof. (This is a fairly standard '3ε argument' in analysis.)

Let $\varepsilon > 0$. Then there exists $k > \frac{3}{\varepsilon}$ such that $\|U - U_k\|_1 < \frac{\varepsilon}{3}$ (by the fact that $U_k \to U$ pointwise almost everywhere, and the dominated convergence theorem). Then $\delta_\square(U, U_k) < \frac{\varepsilon}{3}$ as well.

And there also exists n_0 such that for each fixed k, we have $\delta_\square(W_{n,k}, U_k) < \frac{\varepsilon}{3}$ for all $n > n_0$. (This is because the way we constructed U_k was by taking vertical limits — here we've fixed k and we're looking at a specific vertical column. It converges, so it must be an $\frac{\varepsilon}{3}$-approximation after some point onwards.)

And finally, we know that $\delta_\square(W_n, W_{n,k}) < \frac{1}{k} < \frac{\varepsilon}{3}$ (by the regularity assumption).

Putting these three inequalities together, we find that $\delta_\square(U, W_n) < \varepsilon$. Since $\varepsilon > 0$ can be arbitrarily small, we see that $W_n \to U$ in $\delta_\square$. □

And that finishes the proof of the compactness of the space of graphons.

§15.2 Applications of compactness

Knowing compactness gives us a lot of power; we'll now see how to use it. First we'll use it to prove the existence of limits for a left-convergent sequence.

Proposition 15.13

Let $W_1, W_2, \ldots$ be a sequence of graphons such that $t(F, W_n)$ converges as $n \to \infty$, for every graph F. Then there exists a graphon W such that $t(F, W_n) \to t(F, W)$ as $n \to \infty$.

This is a nice statement because we don't even need to define cut norm to state it.

Proof. Since $(\mathcal{W}_0, \delta_\square)$ i sgraphon, there exists a graphon W (which is the one we'll use) such that along some subsequence W_{n_i}, we have $\delta_\square(W_{n_i}, W) \to 0$ as $i \to \infty$. Then by the counting lemma, if the cut distances converge then the F-densities necessarily converge, so we have $t(F, W_{n_i}) \to t(F, W)$ as $i \to \infty$.

So then W has the right limit along the *subsequence*. But we claim that now we're done — if $t(F, W_{n_i}) \to t(F, W)$ and we know that the original sequence converged, then it must converge to the same number. So that finishes the proof. □

So this was a very quick application of compactness to a statement that didn't even originally concern the space.

Here's another application of compactness. Recall that compactness means that every open cover has a finite subcover (this is the definition you first learn in a topology class). We'll use this to prove the following nice fact.

Proposition 15.14

For every $\varepsilon > 0$, there exists N only depending on ε such that every graphon lies within cut distance ε of some graph on at most N vertices.

A couple of lectures ago, we proved a simpler version of this fact — that every graphon lies within cut distance ε of *some* graph. But the number of vertices in that graph could depend on the graphon, not just ε; so this is a stronger statement.

Proof. For every graph G, we can define the ε-ball around G as

$$B_\varepsilon(G) = \{W \mid \delta_\square(G, W) < \varepsilon\}.$$

(This is some open ball in the metric space.) By the previously proven theorem, every graphon lies within ε cut distance of *some* graph (although potentially with a lot of vertices), so in particular, $\mathcal{W}_0$ is covered by all these balls $B_\varepsilon(G)$ as we range over all graphs. So this gives us an (infinite) open cover, and by compactness this infinite cover has a finite subcover. Then we can let N be the maximum number of vertices among the graphs G in this finite subcover, and we're done. □

Remark 15.15. This proof has some really funny features. The next question you may ask is, what is N as a function of ε? This proof doesn't tell you *anything* about how N depends on ε. You *could* open up the proof of compactness or find a different proof going through similar ideas, and extract a bound; it turns out to be roughly exponential (whatever comes out of the regularity lemma). But if we only knew the compactness statement as a black box, we wouldn't get anything.

In general, this comes up in serious research — proofs using ergodic theory or nonstandard analysis often don't give quantitative bounds, since they end up using things similar to compactness. In the textbook there's a nice theorem that has a proof a few lines long, where we don't know *any* quantitative bounds. It's a very simple statement whose proof is almost trivial, but for which we don't know any function we can write down and prove the theorem for.

§15.3 Equivalence of convergence

Finally, we'll prove the equivalence of convergence (or at least sketch the ideas) — that we have left-convergence (convergence in F-densities) if and only if we have $\delta_\square$-convergence.

The backwards direction was easy — $\delta_\square$-convergence implies left-convergence by the counting lemma. The other direction is more difficult, and much more intricate.

We need to show that if W_1, W_2, ... is a sequence such that $t(F, W_n)$ converges for every F, then we want to show that W_n is a Cauchy sequence with respect to $\delta_\square$.

By compactness, there is always some sequential limit point with respect to $\delta_\square$; call that limit point W. We want to show that this limit point is unique — what would it mean for the conclusion to fail? It's a compact space, so if the sequence were not Cauchy, then there would be two limit points.

So suppose that U is another limit point. (This is kind of abstract nonsense that we'll get to play with in the homework problem — we have a statement about compactness, come up with a hypothetical counterexample, and think about what happens under the limit.) Then by the counting lemma, some subsequence of $t(F, W_n)$ converges to $t(F, W)$, and some other subsequence of the same sequence converges to $t(F, U)$. But we knew from the beginning that the F-densities converge, so these two limits must be equal to each other — in other words, $t(F, W) = t(F, U)$ (for all F).

So it remains to show the following, which is a statement about the uniqueness of F-densities.

Theorem 15.16 (Uniqueness of moments)
If U and W are graphons such that $t(F, W) = t(F, U)$ for all graphs F, then $\delta_\square(U, W) = 0$.

The reason why we call it the uniqueness of *moments* is there's an analogous statement in probability — given a real random variable X, its *moments* are defined as $\mathbb{E}[X^k]$. Suppose we're given all the moment data; does this identify the probability distribution of X? This is a very important problem because in a lot of ways, the way you show convergence to some limit (e.g. the central limit theorme) is by checking the moments — if all the moments converge to the one you want, then you claim this distribution is indeed asymptotically e.g. normal. But that wouldn't work if there existed two different probability distributions with the same set of moments. So this is a fundamental question. And under nice enough hypotheses (the

moments don't blow up too quickly), uniqueness of moments is indeed true. (There are very bad examples — there are nasty examples of distinct distributions sharing moments; but these don't come up with the usual applications.)

Likewise, here the F-densities are some statistic not too dissimilar from kth moments; and this says that if all F-densities agree, then the two graphons must agree with each other.

The proof of this is surprisingly intricate. The rough strategy is the following. We're going to start with W, and considre a couple of finite samples of W. We think of W as a graphon; and we pick $x_1, \dots, x_k$ uniformly in $[0,1]$. We can imagine plotting them on the two axes, and sampling the $k \times k$ set of points in $[0,1]^2$ that they determine. And from that, we can extract $\mathbf{H}(k,w)$; we can think of this as a matrix, or another graphon, which is a $k \times k$ graphon with evenly segmented cells consisting of the values $W(x_i, x_j)$ — the values obtained by sampling each point. (So this is in some sense a $k \times k$ matrix of the sampled values.) We can view this as a graphon, or as a k-vertex edge-weighted graph whose ijth edge weight is given by $W(x_i, x_j)$.

And then from $\mathbf{H}(k,W)$ we can do one more sampling to obtain $\mathbf{G}(k,W)$, which is obtained by keeping edge ij with probability $W(x_i, x_j)$.

So we start with W and pick random vertices; and we look at them and read out the values of W, which are some numbers in $[0,1]$. Then we flip a biased coin to see whether we keep that as an edge or throw it out.

We also zero out the diagonal entries, since we don't want any loops.

> **Remark 15.17.** We don't care about measure-zero sets in W; but if we view $\mathbf{G}$ and $\mathbf{H}$ as probability distributions, then they're well-defined up to probabilitiy 0.

We can couple these processes — from W to $\mathbf{H}$ to $\mathbf{G}$ — by picking $x_1, \dots, x_k$ from the get-go.

The proof strategy is the following. The probability that $\mathbf{G}(k,W)$ is equal to some graph F as *labelled* graphs is a statement about induced subgraph densities — it's completely determined by $t(F', W)$ for all F' which have $v(F) = v(F')$. (If you know all the subgraph densities in W, then you can recover these probabilities. For example, the probabilitiy that $\mathbf{G}(3,W)$ is a triangle is the triangle density; the probability it's a 2-edge path is the density of this path minus the triangle density, with appropriate coefficients.)

So given that U and W have the same F-densities, we see that $\mathbf{G}(k,W)$ and $\mathbf{G}(k,U)$ have the same distribution — theyr'e both random graphs, and they have the same probability distribution (because that's completely determined by the homomorphism densities, which are identical).

The next step is to show that the difference between $\mathbf{G}$ and $\mathbf{H}$ isn't that large in cut metric (i.e., the $\delta_\square$ distance goes to 0 as $k \to \infty$). It's essentially analogous to the difference between $G(n,p)$ and the constant-p graphon — you can use concentration inequalities on each box.

And finally, the process of sampling $\mathbf{H}$ from W is also not very lossy. This also requires some arguments; but if you imagine that W began as a step graphon (i.e., it was very blocky), then by going from W to $\mathbf{H}(k,W)$ with large k is not doing much. And everything can be approximated by step graphons, so this is also true. (This is in fact a L^1-approximation.)

Similarly we have $\mathbf{G}(k,U) \approx \mathbf{H}(k,U) \approx U$; and by taking k large enough, we see that W and U get closer and closer in cut distance. So we can make them arbitrarily close by taking k large enough; this proves what we want.

So this finishes the proof of uniqueness of moments — if U and W agree on all F-densities, then they have zero cut distance. This in turn implies the equivalence of convergence.

This is a really nice proof because the result is actually really nontrivial — it was originally proved not via compactness, but via regularity. Think about a more quantitative statement in the harder direction; that would be what's called an *inverse counting lemma*. We'll state it, and it'll be one of our homework problems to prove it. The counting lemma says that two graphons close in cut norm have similar F-densities.

Theorem 15.18 (Inverse counting lemma)
For every $\varepsilon > 0$, there exists $\eta > 0$ and k such that if $|t(F,U) - t(F,W)| \le \eta$ for all F with $v(F) \le k$, then $\delta_\square(U,W) \le \varepsilon$.

This was first proven quantitatively using regularity techniques; this is a more quantified version of what we proved, and you can actually specify what values η and k need (whereas the compactness proof doesn't give quantiative bounds). On the problem set we'll deduce this from compactness and what we've seen.

§16 November 6, 2023

§16.1 Graph homomorphism inequalities

We'll spend a couple of lectures talking about graph homomorphism inequalities — specifically, *linear* equations between homomorphism densities.

Question 16.1. Given graphs $F_1, \dots, F_k$ (which we think of as fixed), as well as fixed real numbers $c_1, \dots, c_k$, does the inequality

$$\sum_{i=1}^{k} c_i t(F_i, G) \ge 0$$

hold for all graphs G?

This an example of a (linear) graph homomorphism inequality. We'll look at problems like this, as well as variations.

Definition 16.2. The *F-density of G* is

$$t(F,G) = \frac{\hom(F,G)}{v(G)^{v(F)}}.$$

In other words, it's the probability that a random map from $V(F)$ to $V(G)$ is a homomorphism. This has a nice interpretation in the world of graphons; we'll use this language often in this chapter.

§16.2 Some remarks

First, we don't lose much generality by only considering linear inequalities, because we can write

$$t(F_1,G)t(F_2,G) = t(F_1 \sqcup F_2, G).$$

(But it's sometimes still convenient to write polynomials.)

Some of the problems we saw in the first chapter can be interpreted in terms of homomorphism density — Turán's theorem is about maximizing $t(K_2,G)$ subject to the condition $t(K_r,G) = 0$ (what is the maximum edge density in a K_r-free graph)? We gave an answer to this question; we'll now see it again, and next lecture we'll give another proof of Turán's theorem in this format.

This area is quite rich; we'll begin with a few open problems in this line. This question (of whether such an inequality holds) is in general undecidable — there cannot exist a computer algorithm where we feed in such an inequality (with rational coefficients) and the algorithm outputs whether the inequality is true or not true.

It's worth comparing this with related statements. Polynomial inequalities over the *reals* are decidable; this follows from a classic result of Tarski saying that first-order logic over $\mathbb{R}$ is decidable. (Given a polynomial inequality or system of inequalities over $\mathbb{R}$, there is an algorithm to see if it holds. In fact, you can open Mathematica and feed it in, and there are premade libraries that do exactly that, although runtime blows up quickly.)

However, polynomial inequalities over $\mathbb{Z}$ are *undecidable*. We may have seen this in the format that diophantine equations are undecidable. (This was a problem asked by Hilbert.) The case for graphs is a lot more like $\mathbb{Z}$ than $\mathbb{R}$. This may be surprising because graphs look a lot more like $\mathbb{R}$ than $\mathbb{Z}$ in some sense. But in some sense, they're a lot more like $\mathbb{Z}$. And the undecidability of homomorphism inequalities in general should be seen as a sign of richness of this area (similarly to how number theory flourishes because solving diophantines is difficult).

§16.3 Connection to graphons

Recall that the space $(\mathcal{W}_0, \delta_\square)$ of graphons is compact, and $t(F, \cdot)$ is continuous. A continuous inequality is true for all graphs G if and only if it is true for all graphons W. So everywhere above, we can change G to W, and it doesn't affect the substance of the problem.

It's often convenient to have W — a problem that might not have an attainable optimum over graphs will have an attainable optimum over the space of graphons.

For example, consider the problem to minimize $t(C_4, G)$ subject to having $t(K_2, G) \geq p$. The answer to this problem is p^4. This is not attained by any graph, but it is attained by the graphon $W = p$. So instead of asking over the minimum over G, it's much better to ask for the minimum over W.

§16.4 Major open problems

The first open problem we'll talk about is Sidorenko's conjecture.

Definition 16.3. A graph F is *Sidorenko* if $t(F, G) \geq t(K_2, G)^{e(F)}$ for all G.

Conjecture 16.4 (Sidorenko's conjecture) **—** Every bipartite graph is Sidorenko.

In other words, every bipartite graph F satisfies $t(F, G) \geq t(K_2, G)^{e(F)}$ for all G.

One way to interpret this is to phrase it in the optimization form — among all graphs of prescribed edge density $t(K_2, G) = p$, which has the smallest F-density? The conjecture says that when F is bipartite, the minimum is achieved by $W = p$.

Remark 16.5. The 'bipartite' hypothesis is necessary, in the sense that every non-bipartite graph is non-Sidorenko — if F is non-bipartite, then we can take a complete bipartite graph G. Then $t(F, G) = 0$ (we can't embed a non-bipartite graph into a bipartite graph).

Sidorenko's conjecture has been verified for a lot of families of graphs, but it's open for 'almost all' graphs (and some people believe that it is false). The smallest open case is $F = K_{5,5} \setminus C_{10}$, which has the following representation: we can draw 5 vertices on each side, and connect each vertex on the left to the vertex horizontally next to it and the ones immediately above and below. This graph is also known as the *Möbius graph* — we can take a simplicial decomposition of the Möbius strip (where you take a piece of paper and glue two opposite sides in a specific way), i.e., a decomposition into triangles. Then the corresponding graph where we put vertices on one side and faces on the other is exactly this graph.

Another related conjecture — suppose we take F for which Sidorenko's conjecture is true, and ask, what about the equality case? We saw what happens for C_4 in the chapter on quasirandom graphs — you have $t(C_4, G) \approx p^4$ if and only if G is quasirandom. Can you replace C_4 by something else? We say a graph is *forcing* if it can play the role of C_4.

Definition 16.6. A graph F is *forcing* if every graphon W with $t(F, W) = t(K_2, W)^{e(F)}$ is a constant graphon (almost everywhere).

By *constant graphon*, we allow changes up to measure 0. This definition basically says that in the Sidorenko inequality, equality only holds for a constant graphon (and nothign else).

This is related to quasirandomness in the following way.

Proposition 16.7

A grpah F is forcing if and only if for every constant $p \in [0, 1]$, every sequence of graphs $G = G_n$ with $t(K_2, G) = p + o(1)$ and $t(F, G) = p^{e(F)} + o(1)$ is quasirandom.

In other words, C_4 (in the quasirandom equivalences) can be replaced by F if and only if F is forcing.

Question 16.8. Which graphs are forcing?

We saw C_4 is forcing, but what about other graphs?

Claim 16.9 — If F is forcing, then F is Sidorenko.

Proof. If not, then $t(F, G)$ could be both above and below the random-like quantity, and this gives you lots of room to interpolate and get W achieving equality — more precisely, there exist W_0 and W_1 such that $t(F, W_0) > t(K_2, W_0)^{e(F)}$ and $t(F, W_1) > t(K_2, W_1)^{e(F)}$. Then you can interpolate to get a nonconstant W with $t(F, W) = t(K_2, W)^{e(F)}$. (We're being somewhat vague about how you interpolate; but if these are true then W_0 and W_1 are nonconstant, so you can rearrange them so that you don't accidentally hit a constant when interpolating, and then do a linear interpolation.) □

So if you're forcing, you're necessarily Sidorenko. But the converse isn't true. For example, if F is a tree, can it be forcing? As a simple example, if F is a single edge, then it's certainly not forcing — the condition $t(F, W) = t(K_2, W)^{e(F)}$ tells you nothing, so it certainly doesn't imply W is constant. More generally, if F is a tree, then any regular tree satisfies this equality.

The conjecture is that these are the only problematic cases.

Conjecture 16.10 (Forcing conjecture) **—** A graph F is forcing if and only if it is bipartite and has at least one cycle.

These two conjectures are related (though the forcing conjecture has been verified for only a smaller class of graphs).

These are a couple of major open problems in this area, both of which Prof. Zhao likes very much; they are shown here to illustrate the richness of the subject.

Now we turn to some more basic problems.

§16.5 Edge vs. triangle densities

Question 16.11. What are all the possible edge vs. triangle densities?

Definition 16.12. The *edge-triangle region* is defined as

$$\{(t(K_2, W), t(K_3, W)) \mid W \text{ a graphon}\} \subseteq [0,1]^2.$$

This is some closed subset of the unit square, since the space of graphons is compact and F-density is a continuous map. (If we used graphs instead, then we'd get a fractal-like picture with lots of holes.)

This is also related to the following extremal problem:

Question 16.13. Given $t(K_2, W)$, what are the maximum and minimum possible $t(K_3, W)$?

The two problems are actually equivalent (that equivalence has a bit of content in it, as we'll see soon).

We have some region, which is supposed to be the answer to this question. We're saying that if we determine the maximum and minimum, then we've determined the whole region. That means the region is vertically connected. And this can again be shown by interpolation — given W_0 and W_1, we can linearly interpolate between them and continuously go through the densities between them.

§16.6 Maximum triangle density

Proposition 16.14
We have $t(K_3, G) \le t(K_2, G)^{3/2}$.

This is tight, at least asymptotically — if we take G to be a clique (which asymptotically corresponds to W being a graphon taking value 1 in some $a \times a$ box and 0 outside that box), then $t(K_3, W) = a^3$ and $t(K_2, W) = a^2$. (We're being cavalier in using G sometimes and W sometimes, but it doesn't really matter.)

Proof. It's equivalent to prove that $\hom(C_3, G) \le \hom(K_2, G)^{3/2}$ (if we look at how these quantities are defined — there are denominators of n^3 on both sides). We also know that $\hom(C_3, G) = \operatorname{tr} A_G^3 = \sum \lambda_i^3$, and that $\hom(K_2, G) = \operatorname{tr} A_G^2 = \sum \lambda_i^2$. So it remains to verify the inequality

$$\sum \lambda_i^3 \le (\sum \lambda_i^2)^{3/2}.$$

To prove this, we can divide the two sides, to get that

$$\frac{\text{LHS}}{\text{RHS}} = \sum \left(\frac{\lambda_i^2}{\lambda_1^2 + \cdots + \lambda_n^2} \right)^{3/2}.$$

But each fraction is some quantity in $[0,1]$, so we can eliminate the exponent; then

$$\sum \left(\frac{\lambda_i^2}{\lambda_1^2 + \cdots + \lambda_n^2} \right)^{3/2} \le \sum \frac{\lambda_i^2}{\lambda_1^2 + \cdots + \lambda_n^2} = 1. \qquad \square$$

This is a very short proof, but it's also a bit unsatisfying, in the sense that it's a spectral proof of a physical statement (the original statement has nothing to do with eigenvalues). We'll see another proof that's purely physical. It's nice to have physical proofs because they're more versatile — if you start with a non-cycle F then you can't express F-densities in terms of the spectrum, but you still might want to prove quantities.

As a small extension, what happens if we go from triangles to cliques? This turns out to not be too hard.

Proposition 16.15
For all $k \geq 3$, we have $t(K_k, W) \leq t(K_2, W)^{k/2}$.

If we start with this inequality and try to apply the proof method above, we'll run into trouble — there's no good way to express $t(K_k, W)$ in terms of the spectrum. But here we don't need to do that — we can lose a lot and still be okay.

Proof. There exist integers $a, b \geq 0$ such that $k = 3a + 2b$. We can then take K_k and drop a bunch of edges — we just keep a copies of K_3 and b copies of K_2. Then

$$t(K_k, W) \leq t(aK_3 + bK_2, W)$$

(losing edges can only make the subgraph go up, not down — we're dropping factors in the integral). This gives

$$t(K_k, W) \leq t(K_3, W)^a t(K_2, W)^b.$$

And now we can apply the theorem that we just proved to get that this is at most

$$t(K_2, W)^{a/2} t(K_2, W)^b = t(K_2, W)^{k/2}. \qquad \square$$

So this tells us that asymptotically, among graphs with given edge-density, the K_k-density is maximized by a clique. In fact, Kruskal–Katona gives much more precise information — it states that for every number of edges, you maximize the number of k-cliques by clumping together the edges (putting them together into as big cliques as you can). Informally, the number of K_k's is maximized by a clique. If there's some leftover after you form the clique, Kruskal–Katona tells you *exactly* what you should do (putting in the extra edges in a greedy manner).

§16.7 Minimum triangle density

So this solves the problem of maximizing triangle density; the problem of *minimizing* triangle density turns out to be a lot more intricate.

Question 16.16. Given that $t(K_2, W) = p$, how do we minimize triangle density?

If $p \leq \frac{1}{2}$, there is an easy answer — take a bipartite graph, which has triangle density 0.

We're trying to obtain the edge-triangle region as a subset of $[0,1]^2$; we saw an upper boundary of $y = x^{3/2}$, and we've seen that there's a lower boundary up to $x = \frac{1}{2}$ coming from bipartite graphs. What happens after that? This turns out to not be so easy. It turns out that for $p > \frac{1}{2}$, the triangle density is minimized by certain complete multipartite graphs. We'll state the answers without proofs (the proofs are quite difficult — this wasn't known until 15 years ago).

It's useful to have a few reference points. If you take G to be a k-clique, then $(t(K_2, G), t(K_3, G)) = (1 - \frac{1}{k}, (1 - \frac{1}{k})(1 - \frac{2}{k}))$. This gives us a bunch of points with $x = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots$; this turns out to be the best you can do for these specific edge densities (and we will prove this later — that this is optimal for $p = 1 - \frac{1}{k}$ for $k \in \mathbb{Z}$). What happens in between is a lot more intricate. It turns out that the best one can do is the following procedure: for concreteness, let's say that $1 - \frac{1}{3} < p < 1 - \frac{1}{4}$. If $p = 1 - \frac{1}{4}$ then we should take 4 equal parts, and put 1's off the diagonal and 0's on the diagonal. If p is a bit less, we need to find some way to reduce the edge density while minimizing the impact on triangle density. The thing to do is to shrink one of the four parts — and we choose to shrink it just enough to achieve the desired edge density. (As we shrink it, the edge density goes from $1 - \frac{1}{4}$ to $1 - \frac{1}{3}$; at some point it reaches the density that we want.) That this construction is optimal is a difficult theorem.

Theorem 16.17 (Razborov)
Fix $p \in [0,1]$ and let $k = \lceil 1/(1-p) \rceil$. The minimum $t(K_3, W)$ among graphons W with $t(K_2, W) = p$ is attained by the step function with widths $a_1, \ldots, a_k$ and with 0's on the diagonal, 1's off the diagonal, such that $a_1 = \cdots = a_{k-1} \geq a_k$ and $t(K_2, W) = p$.

In graph form, this corresponds to starting with a complete 4-partite graph (or k-partite graph) and then shrinking one of the parts so that the edge density is as given; it turns out that this graph has the smallest triangle density.

Remark 16.18. What does this look like in our plot? We have points corresponding to various cliques on the lower boundary, and connecting them are *concave* curves ('scallops').

Razborov invented a machinery called *flag algebras* to prove this; we'll talk about this next lecture. Part of the difficulty is that these graphs are not unique. For example, we could instead replace the bipartite graph between two of our four graphs by an arbitrary one with the right number of edges; this wouldn't change the number of triangles. Having a large family of equality sets often makes optimization problems more difficult.

A similar story holds when we replace triangles by any r-clique.

§16.8 Cauchy–Schwarz

Now we'll get more down to earth and talk about an important technique for proving such homomorphism inequalities; this is the technique of Cauchy–Schwarz inequalities.

Here's a claim that we've seen before (in the chapter on quasirandom graphs, regarding C_4), but we'll see it again to illustrate notation and get us warmed up.

Theorem 16.19 ($K_{2,2}$ is Sidorenko)
For all W, we have $t(K_{2,2}, W) \geq t(K_2, W)^4$.

We'll prove this claim in two steps.

Lemma 16.20
We have $t(K_{1,2}, W) \geq t(K_2, W)^2$.

We can write $t(K_{1,2}, W)$ as

$$\int_{x,y,y'} W(x,y)W(x,y').$$

(We'll use this notation — where we drop the fact that we integrate over $[0,1]$.)

Note that there's symmetry between y and y', which allows us to put them inside a square — this is equal to

$$\int_x \left(\int_y W(x,y) \right)^2.$$

And now we can apply Cauchy–Schwarz, which lets us pull a square from inside an integral to the outside — this gives that this is at least

$$\left(\int_{x,y} W(x,y) \right)^2 = t(K_2, W)^2.$$

The second step is an elaboration of what we just did.

Lemma 16.21
We have $t(K_{2,2}, W) \geq t(K_{1,2}, W)^2$.

Proof. We can imagine $K_{2,2}$ as with x and x' on the left and y and y' on the right. We can fold y and y' into a square, and write this as

$$\int_{x,x'} \left(\int_y W(x,y)W(x',y) \right)^2 \geq \left(\int_{x,x',y} W(x,y)W(x',y) \right)^2. \qquad \square$$

Visually, the idea is that y was duplicated into a square, which we could take out using Cauchy–Schwarz to get the inequality that we want.

We'll now prove another inequality (from PS1):

Theorem 16.22 (Triangle is common)
We have $t(K_3, W) + t(K_3, 1-W) \geq \frac{1}{4}$.

(On PS1 we showed that if we red-blue color the complete graph, then the monochromatic triangle density is at least $\frac{1}{4}$.)

Proof. We can expand the second term as

$$\begin{aligned} t(K_3, 1-W) &= \int_{x,y,z} (1-W(x,y))(1-W(x,z))(1-W(y,z)) \\ &= 1 - 3 \cdot t(K_2, W) + 3 \cdot t(K_{1,2}, W) - t(K_3, W). \end{aligned}$$

If we add back in $t(K_3, W)$, it cancels out the last term; so the left-hand side of our original expression is $1 - 3t(K_2, W) + 3t(K_{1,2}, W)$. And we saw that $t(K_{1,2}, W) \geq t(K_2, W)^2$. So we now have some elementary quadratic, which we can write as a sum of squares as

$$\frac{1}{4} + 3\left(t(K_2, W) - \frac{1}{2} \right)^2 \geq \frac{1}{4}. \qquad \square$$

We've introduced the word *common*:

Definition 16.23. A graph is *common* if $t(F, W) + t(F, 1-W) \geq 2^{-e(F)+1}$.

Here the right-hand side corresponds to what happens when $W = \frac{1}{2}$ (which is supposed to achieve equality). So triangles are common. When the concept was brought up, it was thought that maybe all graphs are common. But that is very much false; in fact K_4 is not common. (This is not obvious, and the counterexample is quite nontrivial.) Since then there have been more results proving some graphs are common and others are not; there may not be a clean full characterization.

§16.9 A lower bound on triangle densities

(This is also a problem from the first problem set.)

Proposition 16.24

We have $t(K_3, W) \geq t(K_2, W)(2t(K_2, W) - 1)$.

If we plot this in our picture, it's a parabola, and it links all the lower-bound points corresponding to k-cliques. In other words, this inequality actually is an equality for W corresponding to a clique (i.e., $W = W_{K_k}$).

Proof. Since $0 \leq W \leq 1$, we have that $(1-W(x,z))(1-W(y,z)) \geq 0$ (pointwise). Expanding this expression gives that

$$W(x,z)W(y,z) \geq W(x,z) + W(y,z) - 1.$$

So then

$$t(K_3, W) = \int_{x,y,z} W(x,y)W(x,z)W(y,z) \geq \int_{x,y,z} W(x,y)(W(x,z) + W(y,z) - 1).$$

(This might seem a bit unmotivated; we'll have a chance to play with our own inequalities later.) If we expand, then we get $2t(K_{1,2}, W) - t(K_2, W)$. And again we can lower-bound $2t(K_{1,2}, W) \geq 2t(K_2, W)^2$, which finishes the proof. □

§16.10 Undecidability

At this point, it's useful to revisit the theorem mentioned in the beginning, that homomorphism inequalities among graphs are undecidable. This is related to what we've just been talking about. The first time you see homomorphisms, you might think they're like real polynomial inequalities; but they're actually a lot more like integer inequalities. And in fact, undecidability is proved by reducing from integer polynomial inequalities.

The idea is — imagine we look at this picture, and we impose that we're working in the edge-triangle region and that the inequality $t(K_3, W) \geq t(K_2, W)(2t(K_2, W) - 1)$ is an equality. Then we force our graph to be on this curve, and there's a discrete set of points we're restricted to, corresponding to integers. So we can extract integers in this way — if we start with an integer system of inequalities, we can convert it to a graph homomorphism system of inequalities in this way (encoding our integers via these vertices of the graph). The theorem says a bit more — what we've just said allows you to convert a system of integer inequalities into a system of homomorphism inequalities. In fact the theorem says that you can just use *one* inequality and that'll still be undecidable; there's more work in converting a system of inequalities into one through various gadgets, which we won't discuss. But this picture should show where the diophantineness comes in — because we have these discrete sets of vertices in the graph, which is very different from continuous objects related to polynomial inequalities in the real numbers.

§16.11 Another example

Proposition 16.25

The graph consisting of two four-cycles glued together at an edge is Sidorenko.

Proof. Let this graph be F. We can try using Cauchy–Schwarz — visually Cauchy–Schwarz works by folding a picture in half, and here there's a line of symmetry we can try to fold along. Here's how this works algebraically — let w and x be the two vertices in the middle, and y and z the two vertices on one side and y' and z' on the other. Then we can write

$$t(F, W) = \int_{xw} \left(\int_{yz} W(wy)W(yz)W(zw) \right)^2 W(w,x).$$

(The integral inside the square corresponds to the duplicated vertices and edges; the edge wx isn't duplicated.) Now we do Cauchy–Schwarz. Here we can't just pull the square out, because we have an extra factor. But Cauchy–Schwarz lets us multiply by the extra factor $\int_{wx} W(wx)$ on both sides and *then* use Cauchy–Schwarz, to get that

$$\begin{aligned} t(F,W)t(K_2,W) &= \left(\int_{xw}\left(\int_{yz} W(wy)W(yz)W(zw)\right)^2 W(w,x)\right)\left(\int_{wx} W(wx)\right) \\ &\ge \left(\int_{xwyz} W(wy)W(yz)W(zw)W(wx)\right)^2 = t(C_4,W)^2 \\ &\ge t(K_2,W)^8. \end{aligned}$$

This gives the right answer. (If you do this properly, then you should always get the right exponent in the end.) □

We've seen a number of fairly simple calculations using Cauchy–Schwarz. So there are two questions. Can you always prove what you want using Cauchy–Schwarz? The answer is no — if you could, then you could decide the problem by searching through the space of all Cauchy–Schwarz solutions (contradicting undecidability).

The second is, how do you find such proofs when they do exist? It turns out proofs by Cauchy–Schwarz can be equivalently formulated as proofs expressing some quantity as a sum of squares. One way to prove the Cauchy–Schwarz inequality — which states that $(\int f^2)(\int g^2) \ge (\int fg)^2$ — is to write it as a sum of squares (which shows that it's nonnegative).

We'll elaborate more on this next time, but this forms a space of possible solutions you can search over using a computer (it amounts to semidefinite programming). This has been fruitful in proving some old conjectures, where researchers formulate the right problem and feed it into a semidefinite program, which outputs a very elaborate Cauchy–Schwarz proof that you would not have found by hand. The point is to say that even though the examples we saw today are fairly simple (and can be found by hand), there are examples from research where this method (combined with computation) can take you a lot further.

§17 November 8, 2023

Today we'll talk about additional techiques for proving graph homomorphism inequalities.

§17.1 Cauchy–Schwarz and flag algebras

Last time, we saw a simple application of Cauchy–Schwarz to prove that $t(K_{1,2},W) \ge t(K_2,W)^2$ — we rewrote the left-hand side as $\int_x (\int_y W(x,y))^2$ (when we expand the square we get y and y', giving $K_{1,2}$ — this takes advantage of the fact that the two play symmetric roles). Cauchy–Schwarz allows us to expand this square to get that this is at least $(\int_x \int_y W(x,y))^2$. Last time we commented that Cauchy–Schwarz proofs are always sum-of-squares (SOS) proofs — as a concrete example, here we can write the left-hand side minus the right-hand side as

$$\frac{1}{2}\int_x \left(\int_y W(x,y) - t(K_2,W)\right)^2.$$

(You can check that when you expand the right-hand side, you get the right difference.) This should be reminiscent of the fact that $\operatorname{Var} X = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ is nonnegative, and one way to see this is to rewrite it as $\mathbb{E}[(X - \mathbb{E}X)^2]$. (This is the same calculation, with X being a random variable representing degree.)

Any time you have a Cauchy–Schwarz proof, you can write it as a sum of squares. So it makes sense to talk about this type of proof in a systematic way. This leads us to flag algebras, devised by Razborov in 2007 (and initially used to prove the edge-triangle region — given an edge density, what's the minimum triangle density a graph can have?).

The idea is to tackle problems of the following form, using a specific type of proof that can be computerized (so that you can search for proofs using a computer).

The type of problems this kind of method solves are where we want to minimize or maximize $t(F_0, W)$ (or some linear combination of such graph homomorphism densities), subject to various linear constraints, e.g., $t(F_i, W) = q_i$. (You can imagine more varied forms of this.) Turán's problem is of this form — it asks, if you have zero triangle density, what is the maximum edge density?

We're given these constraints, and we can also introduce 'free' nonnegative constraints on these homomorphism densities. For example, every time you have some expression which just by its formula is clearly nonnegative (e.g. $\int W(x,y)W(x,z)(\int_{u,w} aW(x,u)W(y,u)+\ldots)^2$), no matter what coefficients we have inside (with $a \in \mathbb{R}$), this expression is nonnegative (since we have a square).

Perhaps by writing down enough constraints like this, we'll be able to obtain some bound that we're looking for, simply by manipulating these expressions.

We're trying to prove some inequality — that $t(F_0, W)$ is at least something. And how do you prove inequalities? One is to write that inequality as a sum of squares, and this gives a systematic way to try to do this.

One question is, how do you choos the coefficients? Here we have some indeterminates, and whatever we plug in you get some correct inequality; and we hope that if you put in the right coefficients you get what you want. But how do you choose the right coefficients?

This type of problem turns out to be very well-studied; they're *semidefinite programs*. A *linear program* is where we have linear constraints, and there are fast algorithms. Semidefinite programs are an extension of linear programming, usually given in terms of a semidefinite matrix; usually it's for solving problems where your coefficients come from a bunch of squares. And there are also fast algorithms that work in practice, well enough to have led to some practical successes.

We'll see a few success stories of theorems that people have managed to prove using this method, in combination with an actual computer search (using computers to find the right coefficients, which would be unrealistic to find without computer help).

Theorem 17.1

Every n-vertex triangle-free graph has at most $(n/5)^5$ cycles of length 5.

(This was an Erdős conjecture from the 1980s.)

This asks to maximize the number of 5-cycles in a triangle-free graph; the construction is the blowup of a 5-cycle (where we split the vertices into five equal parts, and put in all the edges going around — this turns out to be the best you can do).

This is a very nice result, that was solved using the flag algebra method. The proof was essentially an explanation of the method, and in the appendix were the matrices with crazy rational coefficients.

Here's another problem, about inducibility.

Theorem 17.2

Every n-vertex graph has at most $\frac{n^5}{5^5-5}$ induced 5-cycles.

Here we don't have other constraints; to get induced, you need a balance of edges and non-edges. The tight construction is an iterated blowup of a 5-cycle — we first blow up a 5-cycle, and then inside each of the five parts we do another blowup of a 5-cycle (equipartition into 5 parts and draw the edges going around), and keep doing this (until we just have 1-vertex parts).

Again, this was conjectured (or at least asked) in the 1970s, and only proved in 2006 using the flag algebra method. There's more involved here than just writing down the right coefficients — even the structure of the answer seems a bit crazy, and you have to do more work. But the flag algebra method is an essential part of the solution.

It's a very powerful method, and has had a lot of successes. In practice, it's kind of hard to get started. In principle, you can come up with what is the computer program you need to write, but in practice it's annoying because you have to enumerate over all graphs of a certain size and look at various combinations, and get a good SDP solver (which is actually not that easy — there are good packages for linear programming, but SDP is iffier). And as a practical note, there aren't very good well-developed packages for doing this sort of thing. Prof. Zhao had a paper where he and his students ended up with a Turán problem, and eventually he caught the interest of one of the flag algebra experts. There might only be 2 or 3 people in the world who have the setup to do this; and they solved the problem.

§17.2 Incompleteness

Question 17.3. Can every graph homomorphism linear inequality be proven using Cauchy–Schwarz or sum of squares?

The answer is very resoundingly no (as stated last time). But before this, here's a story about real polynomials, which is actually quite interesting.

Question 17.4. What about real polynomial inequalities? In other words, suppose $p(x_1, \dots, x_n) \geq 0$ is a polynomial that's always nonnegative (for all real inputs). Can p be written as a sum of squares?

The answer is yes for one-variable polynomials — we can always factor 1-variable polynomials and write them as a sum of squares. It turns out that the answer is also yes for 2 variables, and no for at least 3.

Hilbert proved this, but his proof was abstract — he showed there *exist* polynomials which can't be written as a sum of squares. The first actual example was given later by Mozkin 1967, who showed that

$$p(x, y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2$$

is always nonnegative (by AM–GM) but not a sum of squares. Important progress was made by Emil Artin in 1927, who showed that every nonnegative polynomial can be written as a sum of squares of *rational* functions — if we're allowed to use denominators, then we can always write p as a sum of squares. For example, the above polynomial can be written as a sum of squares by first multiplying by $(x^2+y^2)^2$ — then we can write it as

$$\frac{x^2y^2(x^2+y^2+1)(x^2+y^2-2)^2 + (x^2-y^2)^2}{(x^2+y^2)^2}.$$

Remark 17.5. Why does our counterexample have $n = 2$? The 2 vs. 3 statement may be for homogeneous polynomials.

That's the story for reals. There's more to say — you can ask Mathematica to decompose a polynomial as a sum of squares, or even to give a sum of squares proof to a system of polynomial inequalities. This is computationally quite expensive.

Coming back to graph inequalities, can every graph inequality be written as a sum of squares? The answer is no. Graph inequalities are undecidable. But if the answer were yes, then we could find an algorithm to decide the problem — in parallel we run two algorithms, one checking all graphs, and the other checking all SOS proofs. Either the inequality would have a counterexample or a SOS proof, so this would work. By undecidability this technique cannot work, so not every inequality has a SOS proof.

However, this is an abstract proof that doesn't give you any specific example of an inequality — can you show me a graph inequality which doesn't have a SOS proof? More recent results give us some things that have no SOS proofs. For example, Sidorenko's inequality for the 3-edge path — that $t(P_3, W) \geq t(K_2, W)^3$ — has no SOS proof (this was proven very recently), even if you are allowed to multiply by rational squares. The same paper also shows that Sidorenko for the Möbius graph (which is still open — we don't know if it's true) cannot be proven using SOS.

So this tells us there are limitations to this method — it's a nice proof system, but it has certain limitations.

§17.3 Hölder's inequality

Hölder's inequality is in a way one step up from Cauchy–Schwarz.

Theorem 17.6

For $p_1, \dots, p_k$ such that $\sum \frac{1}{p_i} = 1$, we have

$$\int f_1 \cdots f_k \leq \|f_1\|_{p_1} \cdots \|f_k\|_{p_k},$$

where $\|f\|_p = (\int |f|^p)^{1/p}$.

The case where $p_1 = \cdots = p_k = \frac{1}{k}$ is an important case we'll see quite a lot. For example, last time we showed that $K_{2,2}$ is Sidorenko. Now using the same method but with Hölder's inequality, we can show $K_{s,t}$ is Sidorenko.

Theorem 17.7

$K_{s,t}$ is Sidorenko.

We won't write out all the details, but we can imagine having an inner integral and an outer integral. We first use Hölder's inequality to take the t repeated variables and put them together, writing $t(K_{s,t}, W) \geq t(K_{s,1}, W)^t$; and then we similarly show this is at least $t(K_{1,1}, W)^{st}$, and we're done.

Here's something more interesting:

Theorem 17.8

The 3-edge path is Sidorenko.

This is a deceptively simple-looking statement. The proof is not very long, but it is quite tricky and requires some new ideas beyond the Cauchy–Schwarz type calculations we did earlier.

We'll see a few different proofs, which are all quite nice.

Write the vertices of our 3-edge path as w, x, y, z (in order).

Proof 1. We have $t(P_3, W) = \int_{wxyz} W(xw)W(xy)W(zy)$. We'll replace one of the ends by a degree function

$g(x) = \int_y W(x,y)$ (here $g(x)$ essentially represents the degree at vertex x), so that

$$t(P_3, W) = \int_{xyz} g(x)W(xy)W(zy)$$

(we're essentially integrating out the w variable).

We can also do the same thing to the other end, namely z — so we also have

$$t(P_3, W) = \int_{wxy} W(xy)W(zy)g(z).$$

(Here we chop off the leaf on the other end, replacing its neighbor with its degree.) If we take the geometric mean of these two expressions and apply Cauchy–Schwarz, we get that

$$t(P_3, W) \geq \int_{xyz} \sqrt{g(x)}W(xy)W(zy)\sqrt{g(z)}.$$

Now this expression is symmetric down the middle, so we can write it as

$$\int_y \left(\int_x \sqrt{g(x)}W(xy)\right)^2.$$

And using Cauchy–Schwarz again, this is at least

$$\left(\int_{xy} \sqrt{g(x)}W(xy)\right)^2.$$

But g was a degree function, so we can integrate out the y and replace it by $g(x)$ to get

$$\left(\int_x g(x)^{3/2}\right)^2.$$

So far we've only used Cauchy–Schwarz, but as stated this inequality can't be proven just using Cauchy–Schwarz. Now using Hölder with exponent $\frac{3}{2}$, we get that this is at least $(\int_x g(x))^3 = (\int_{xy} W(x,y))^3$, which is exactly what we wanted. □

Proof 2. Defining g as above, we have $t(P_3, W) = \int_{wxyz} W(xw)W(xy)W(zy) = \int_{xy} g(x)W(xy)g(y)$ (replacing both the head and tail by g). Let's now add in a couple of extra terms that are both equal to 1 — we write this as

$$\left(\int_{xy} g(x)W(xy)g(y)\right)\left(\int_{xy} \frac{W(x,y)}{g(x)}\right)\left(\int_{xy} \frac{W(x,y)}{g(y)}\right).$$

(For the second term, if we integrate over y we get $g(x)$, so the term is 1; the same is true for the third.) Now applying Hölder's inequality this is at least $(\int_{xy} W(xy))^3$, and we're done. □

§17.4 A generalization of Hölder's inequality

Here's one specific version:

Theorem 17.9

Suppose that we have $f\colon X \times Y \to \mathbb{R}$, $g\colon X \to \mathbb{Z} \to \mathbb{R}$, and $h\colon Y \times Z \to \mathbb{R}$. Then

$$\int f(xy)g(xz)h(yz) \leq \|f\|_2 \|g\|_2 \|h\|_2.$$

Vanilla Hölder would give something related to 3-norms (though we have to be careful because there's a missing variable). But there's a better inequality — we can instead get 2-norms. (This is better because in probability spaces, the common setting, we have monotonicity of norms.)

Proof. We essentially apply Cauchy–Schwarz three times, once to each variable. We start with the LHS. Now we do Cauchy–Schwarz on the variable x; when we do this, we can ignore the third factor, so we get that

$$\text{LHS} \le \int_{yz} \left(\int_x f(xy)^2 \right)^{1/2} \left(\int_x g(xz)^2 \right)^{1/2} h(yz).$$

Now we do Cauchy–Schwarz on y. The first and third factors involve y, so we apply Cauchy–Schwarz to those two and get

$$\int_z \left(\int_{xy} f(xy)^2 \right)^{1/2} \left(\int_x g(xy)^2 \right)^{1/2} \left(\int_y h(yz)^2 \right)^{1/2}.$$

And finally we apply Cauchy—Schwarz to the last remaining variable z; we apply it just to the second and third terms, and then we get $\|f\|_2 \|g\|_2 \|h\|_2$. □

Remark 17.10. Would this let us prove the maximum triangle density? Yes — setting $f = g = h = W$, this tells us that

$$t(K_3, W) \le \left(\int W^2 \right)^{3/2}.$$

But W is a graphon, so it has values in $[0, 1]$, and we can bound this by $(\int W)^{3/2}$.

There, we complained that a purely physical inequality had a proof that had to go through the spectrum — there's nothing wrong with this, but it's nice to also have a physical proof, and this is one.

It also gives us more — it gives us an intermediate step, where we have $\int W^2$. For the purpose of usual homomorphism inequalities we should ignore it, since it doesn't have any meaning. It's equivalent to talk about graphs and graphons, and for graphs W^2 and W are identically equal.

But this is a true inequality, and it still makes sense for graphons taking intermediate values. And there are cases where this is important — Prof. Zhao used it when working on upper tails of subgraphs counts (where there's a nonlinear inequality, and this matters).

Remark 17.11. In this proof we should first replace f, g, and h by their absolute values (to prevent weird things from happening).

One way to interpret this inequality — or at least a special case of it — is as a *projection inequality*.

Question 17.12. Given a body (compact set) $K \subseteq \mathbb{R}^3$, if we know the projection of K on each coordinate plane has area at most 1, what is the maximum possible volume of K?

So we see the areas of the three shadows; can we conclude something about its volume? For example, if K is the unit cube, then it has area-1 projection on each coordinate plane and volume 1. And it turns out that this is the best — we have $\text{Vol}(K) \le 1$.

This is also an approximate form of an isoperimetric inequality — these projection areas form a proxy for surface area (up to a constant factor), so this tells us that a bound on the surface area gives a bound on the volume.

More generally, we have the following:

Theorem 17.13

We have $(\operatorname{Vol} K)^2 \le \operatorname{Area}(\pi_{xy}K)\operatorname{area}(\pi_{xz}K)\operatorname{area}(\pi_{yz}K)$.

Proof. Take the function $\mathbf{1}_K$; this is at most the product of the three shadow indicators, i.e.,

$$\mathbf{1}_K \le \mathbf{1}_{\pi_{xy}K}(xy)\mathbf{1}_{\pi_{xz}K}(xz)\mathbf{1}_{\pi_{yz}K}(yz).$$

Then applying the generalization of Hölder gives the desired inequality. □

Using the same method, we can prove a more general inequality.

Theorem 17.14

Suppose we have measure spaces $X_1, \dots, X_m$ and index sets $I_1, \dots, I_\ell$ such that each element of $[m]$ appears in exactly k different I_i's, and suppose that $f_i \colon \prod_{j\in I_i} X_j \to \mathbb{R}$. Then we have

$$\int_{X_1\times\cdots\times X_m} f_1(\pi_{I_1}(x))\cdots f_\ell(\pi_{I_\ell}(x)) \le \|f_1\|_k \cdots \|f_\ell\|_k .$$

The reason we have 2 in the original inequality is that x, y, and z each appear in 2 of the functions; and if we replace 2 by k, the same holds.

Corollary 17.15

If F has maximum degree at most k, then

$$t(F,W) \le \|W\|_k^{e(F)} \le t(K_2,W)^{e(F)/k}.$$

§17.5 Independent sets

Another application of these projection inequalities concerns problems where we are talking about some statistical quantities on graphs, such as the number of independent sets.

Definition 17.16. Let $i(G)$ be the number of independent sets of G.

Theorem 17.17 (Kahn, Zhao)

For every n-vertex d-regular graph G, we have

$$i(G)^{1/n} \ge i(K_{d,d})^{1/2d}.$$

So the left-hand side is maximized by $K_{d,d}$. The connection is that $i(G)$ can be viewed as the number of homomorphisms from G to a 2-vertex graph with one edge and one self-loop (since the vertices mapped to the loopless vertex can't have any edges).

Kahn's original proof (for bipartite graphs) uses entropy; we'll see a different proof using Hölder.

Theorem 17.18

For every d-regular *bipartite* F, we have $t(F,W) \le t(K_{d,d},W)^{e(F)/d^2}$.

By taking W to be the graphon associated to this graph (it turns out that we don't lose anything from the normalization, since the denominators cancel out), we get the desired inequality.

Proof. We'll prove this for F a 6-cycle (since the ideas are the same); we'll label the vertices x_1, x_2, x_3 and y_1, y_2, y_3. Let $f(x_1, x_2) = \int_y W(x_1, y)W(x_2, y)$, which represents the codegree of x_1 and x_2.

We can write $t(C_6, W)$ as an integral of 6 different factors, but we can group them according to x_1, x_2, and x_3 — we can write this as

$$t(C_6, W) = \int_{x_1,x_2,x_3} f(x_1, x_2)f(x_1, x_3)f(x_2, x_3)$$

(where the first term accounts for the top two edges, and so on). By Hölder's inequality this is at most $\|f\|_2^3$. But we have

$$\|f\|_2^2 = \int f(x_1, x_2)^2 = t(C_4, W)$$

(since we're integrating the codegree squared, and that's precisely the C_4-density). And that's it. □

(The same proof, just with a bit more notation, proves this theorem in general.)

There's a lot more to say about this type of problem, including more recent results where they had to use much more involved Hölder-type inequalities to prove an analogous statement for colorings.

§17.6 Lagrangians

We'll now discuss another method, unrelated to what we've seen before. We'll first use this method to give another proof of Turán's theorem.

Theorem 17.19

Every K_{r+1}-free graph has at most $(1 - \frac{1}{r})\frac{n^2}{2}$ edges.

We proved something slightly stronger about the exact number, but this is good enough for most purposes.

The method of Lagrangians is to do some relaxation by allowing the vertices to have weights.

Proof. Let G be a K_{r+1}-free graph on vertex set $[n]$. Let's consider the function

$$f(x_1, \dots, x_n) = \sum_{ij \in E(G)} x_i x_j.$$

(This is supposed to record what happens if we assign node weights to the vertices of G — imagine assigning weights $x_1, \dots, x_n$, and now trying to track weighted edges.) We want to show that $f(\frac{1}{n}, \dots, \frac{1}{n}) \le \frac{1}{2}(1 - \frac{1}{r})$, which corresponds to what happens for a r-clique. In fact, we'll show something slightly stronger — we'll relax the problem to consider all possible node weights that sum to 1. (Initially we start with equal node weights, but we'll allow ourselves to choose different weights, as long as they sum to 1.) And we'll show that no matter what happens, f is at most the claimed quantity — we consider all $x_i \ge 0$ with $x_1 + \cdots + x_n = 1$, and we want to show that $\max f(x_1, \dots, x_n)$ is at most $\frac{1}{2}(1 - \frac{1}{r})$.

This is a compact domain, so by compactness the maximum is attained at some $(x_1, \dots, x_n)$. There could be multiple maxima; we choose one with minimal possible support size (i.e., the fewest nonzero entries). Once we've done this, we can forget about the zero coordinates.

Suppose there exists some non-edge of G with both node weights positive — i.e., $ij \in E(G)$ with $x_i, x_j > 0$. Let's imagine replacing $(x_i, x_j) \rightsquigarrow (x_i + s, x_j - s)$. We know f is a quadratic polynomial; but if we only

perform this change, and there's no edge x_ix_j, then f changes linearly in s. But since we're at a maximum, it has to be constant. So if we're trying to do maximization, we can simply pull s to one endpoint or another — by replacing (x_i, x_j) with $(x_i + x_j, 0)$ we can lose a coordinate, contradicting the minimum support hypothesis.

(We could do this by compactness, or by continuously trying to optimize by pushing things around every time you see this setup.)

So then the support of x is a clique in G. And by relabelling vertices, let's suppose that the clique is supported on the first k coordinates — so $x_1, \dots, x_k > 0$ and $x_{k+1} = \cdots = x_n = 0$. We also know that $k \le r$, since G is K_{r+1}-free.

And now we're basically done — we have that

$$f(x) = \sum_{1 \le i < j \le k} x_i x_j.$$

This is a complete elementary symmetric polynomial, and by using your favorite inequalities (e.g. sum of squares), it's at most

$$\frac{1}{2}\left(1 - \frac{1}{k}\right)(x_1 + \cdots + x_k)^2.$$

And $x_1 + \cdots + x_k = 1$ and $k \le r$, so we're done. □

Remark 17.20. Could there be multiple points achieving the maximum? Yes — if you start with a complete r-partite graph, it's achieved on any r-clique.

Finally, we'll see how to decide linear inequalities between clique densities.

Theorem 17.21

Let $c_1, \dots, c_\ell \in \mathbb{R}$. Then the inequality

$$\sum_{r=1}^{\ell} c_r t(K_r, G)$$

is true if and only if it is true for all $G = K_n$.

So if we have a linear inequality where all the graphs involved are cliques, then to check it for all graphs G, it's sufficient to check it for cliques. This is in huge contrast to general classes of inequalities, which are undecidable. (This inequality is actually a polynomial in n, so it's not hard to check — you can check whether a single-variable polynomial in n is always nonnegative.)

This has a lot of nice implications. As one, when we talked about the edge-triangle region, we proved some upper bound, but we didn't prove the lower bound — we said that there were vertices corresponding to when G is a clique. This theorem tells us that the convex hull of the region is contained in that of our vertices (by drawing line segments along these points) — because to check every *linear* inequality you have to check cliques (and checking linear inequalities corresponds to determining a convex hull), and these linear inequalities are completely determined by these nodes.

Corollary 17.22

In $\mathbb{R}^{\ell-1}$, the convex hull of $\{(t(K_2, W), t(K_3, W), \dots, t(K_\ell, W))\}$ (for all graphons W) is the convex hull of the same set, but just for $W = W_{K_n}$.

Proof. The only nontrivial part is the 'if' implication (we want to show that if the inequality is true for all cliques, then it's true for all graphs).

Suppose it's true for all cliques G. Now let G be an arbitrary graph with vertex set $[n]$, and let

$$f(x_1, \dots, x_n) = \sum_{r=1}^{\ell} r!\, c_r \sum_{r\text{-cliques}} x_{i_1} \cdots x_{i_r}$$

(we sum over r-cliques in G and take products of their node weights). The idea is similar to before — we relax to allow node weights in G, and this is then the quantity corresponding to our expression. (The expression we're interested in is $f(\frac{1}{n}, \dots, \frac{1}{n})$.) We want to show that this quantity is nonnegative. We'll show something much stronger — that for *every* choice of ode weights $x_1, \dots, x_n \geq 0$ with $x_1 + \cdots + x_n = 1$, we have $f(x_1, \dots, x_n) \geq 0$.

The idea is that we have this inequality in some real variables, and we want to be able to shift some weights — we look at two vertices and see if we can shift weight from one to the other. We keep doing this until we can get down to most coordinates being 0, which gets us back down to a clique situation.

By compactness, the minimum is attained at some x; choose one with minimum support. If $ij \notin E(G)$ but $x_i, x_j > 0$, then we're going to replace the two node weights (x_i, x_j) with $(s, x_i + x_j - s)$. As before, since there's no edge between them, f is linear in s; and since we're at the minimum, it has to be constant in s. So we can replace one of the two weights by 0, contradicting minimum support size.

Now we're down to a situation where $G = K_n$ (in the sense that we can delete all the other vertices that don't matter anymore). Ad once G is a clique, we have a symmetric polynomial in the inputs.

Lemma 17.23

Let $f(x_1, \dots, x_n)$ be a real linear combination of elementary symmetric polynomials. If $f(x_1, \dots, x_)$ is minimized among all x with $x_1, \dots, x_n \geq 0$ and $x_1 + \cdots + x_n = 0$ and has minimum support among such x's, then up to permuting coordinates, we have $x_1 = \cdots = x_k = \frac{1}{k}$ and $x_{k+1} = \cdots = x_n = 0$.

This specifically refers to symmetric polynomials which have no quadratic terms (i.e., no squares).

And this says that then we might as well be checking a clique (with even node weights).

This is an exercise in algebra; if you are at a minimum, you can still do this argument and think about what happens when you try to move weight. It's some quadratic polynomial, which can either be concave up or concave down. One tells you the two x and y should be equal (here we're using symmetricness), and the other tells you they should be zero. And if among n numbers they have to be pairwise equal or one zero, these are the only possibilities. □

§18 November 13, 2023

We're now going to talk about additive combinatorics; but many themes from the first part of the class will reappear.

In this section we'll discuss Roth's theorem.

Theorem 18.1 (Roth)

Every 3-AP-free subset of $[N]$ has size $o(N)$.

Earlier, we saw a proof using graph theory. Today and next lecture we'll see Roth's original proof, which uses Fourier analysis; this is a central topic in additive combinatorics.

We'll see the proof of Roth's theorem in this form next time; today we'll prove a variant of Roth's theorem in the finite field model, specifically in $\mathbb{F}_3^n$.

Theorem 18.2 (Roth's theorem in $\mathbb{F}_3^n$)
Every subset of $\mathbb{F}_3^n$ that is 3-AP-free has size $o(3^n)$.

Compared to the version of Roth's theorem in $\mathbb{Z}$, we've replaced $\mathbb{Z}$ by a vector space over a finite field. There is no implication between these two statements — they're completely different — but they are analogous (one is about sets in $\mathbb{Z}$, and the other about sets in a finite group).

We'll see the proof of the latter today using Fourier analysis. Most of the ideas will carry over to $\mathbb{Z}$, which we'll do next time. On the other hand, the finite field model is a lot easier to work with, and it's an important starting point to play with many ideas in additive combinatorics — you should see it as a playground to try out ideas which often can be carried over to other settings, but are technically easier here.

So the goal for today is to prove Roth's theorem in $\mathbb{F}_3^n$; on Wednesday we'll transfer the proof idea to the integers.

§18.1 Primer on Fourier analysis

Here's a crash course on Fourier analysis in the ways that are relevant to us.

We're going to talk about the Fourier transform in a finite field vector space. Usually the first place you see Fourier analysis is when looking at periodic functions. The discrete Fourier transform is in some ways simpler because there are no issues with continuity or technical conditions from analysis, but the ideas are similar.

Definition 18.3 (Fourier transform in $\mathbb{F}_p^n$)**.** Given a function $f\colon \mathbb{F}_p^n \to \mathbb{C}$, we define its *Fourier transform* $\widehat{f}\colon \mathbb{F}_p^n \to \mathbb{C}$ by setting $\widehat{f}(r)$, for each $r \in \mathbb{F}_p^n$, as

$$\widehat{f}(r) = \mathbb{E}_{x\in\mathbb{F}_p^n} f(x)\omega^{-r\cdot x},$$

where $\omega = \exp(2\pi i/p)$.

(The exponent of ω being an element of $\mathbb{F}_p$ makes sense, because $\omega^p = 1$; by $\mathbb{E}_{x\in\mathbb{F}_p^n}$ we really mean $\frac{1}{p^n}\sum_{x\in\mathbb{F}_p^n}$, and the dot product is the usual dot product $r\cdot x = r_1x_1 + \cdots + r_nx_n$.)

This should be reminiscent of other forms of the Fourier transform we've seen, such as Fourier coefficients for a periodic function.

Today we'll give proofs of the basic properties of the Fourier transform, and then use them to prove Roth's theorem (in $\mathbb{F}_3^n$).

The definition of the Fourier transform gives a way to obtain $\widehat{f}$ from f. The *Fourier inversion formula* allows us to go backwards.

Theorem 18.4 (Fourier inversion formula)
For any $f\colon \mathbb{F}_p^n \to \mathbb{C}$, we have

$$f(x) = \sum_{r\in\mathbb{F}_p^n} \widehat{f}(r)\omega^{r\cdot x}.$$

This formula looks very similar to the formula for the Fourier transform itself, but with two important differences. One is that instead of an expectation we have a summation; the other is that there is no minus sign in the exponent. (We'll soon give a conceptual explanation for why these changes have to be there.)

Another important property is known in the literature under two names, which are often interchangeable (even though they are totally different people); one is Plancherel, and the other is Parseval.

Theorem 18.5 (Parseval/Plancherel)
Given $f, g: \mathbb{F}_p^n \to \mathbb{C}$, we have

$$\mathbb{E}_{x \in \mathbb{F}_p^n} f(x)\overline{g(x)} = \sum_{r \in \mathbb{F}_p^n} \widehat{f}(r)\overline{\widehat{g}(r)}.$$

In particular (when $f = g$), we have

$$\mathbb{E}_{x \in \mathbb{F}_p^n} |f(x)|^2 = \sum_{r \in \mathbb{F}_p^n} \left|\widehat{f}(r)\right|^2.$$

Remark 18.6. Regarding the naming, historically Parseval discovered the version of this identity for Fourier series (for periodic functions on $\mathbb{R}$), and Plancherel proved a version for the Fourier transform on $\mathbb{R}$. But they're abstractly both the same idea; nowadays the two names are used interchangeably to refer to the unitarity of the Fourier transform by these identities.

Remark 18.7. There may seem to be some asymmetry where sometimes we see an $\mathbb{E}$ and sometimes a $\sum$. There's actually a good reason for this. The modern convention is that we always average in physical space (e.g. $\mathbb{E}f$) and sum in the frequency or Fourier space (e.g. $\sum \widehat{f}$). For example, in Parseval the right-hand side concerns the Fourier transform so we sum, while the left-hand side concerns physical quantities so we take expectations.

This is a good convention to remember, and if you stick with it you generally won't make mistakes.

We've stated a bunch of identities. Their proofs are quite straightforward, in that you can plug in the definitions and expand, and see that everything that should cancel will. This is a lot simpler than what happens in the real analysis setting, where you sometimes have to think about convergence. (If you've seen the Fourier inversion formula for the Fourier transform in $\mathbb{R}$, it's tricky to prove because you can't actually exchange certain integrals.) But here there are no issues — everything is a discrete sum, so things work out easily.

Still, we'll give conceptual proofs of these identities that will help us think about what these quantities are. It's helpful to see that there's something that clearly looks like an inner product, and we can certainly define those.

Definition 18.8. Given two *physical* functions $f, g: \mathbb{F}_p^n \to \mathbb{C}$, we define

$$\langle f, g \rangle = \mathbb{E}_x \overline{f(x)} g(x) \text{ and } \|f\|_2 = \langle f, f \rangle^{1/2}.$$

Similarly, for functions $\alpha, \beta: \mathbb{F}_p^2 \to \mathbb{C}$ in frequency space, we define

$$\langle \alpha, \beta \rangle_{\ell^2} = \sum_x \overline{\alpha(x)}\beta(x) \text{ and } \|\alpha\|_{\ell^2} = \langle \alpha, \alpha \rangle_{\ell^2}^{1/2}.$$

Let $\gamma_r: \mathbb{F}_p^n \to \mathbb{C}$ (for $r \in \mathbb{F}_p^n$) as

$$\gamma_r(x) = \omega^{r \cdot x}.$$

This is known as a *character* (it's a multiplicative function). The Fourier transform can then be written in terms of characters, as

$$\widehat{f}(r) = \mathbb{E}_x \overline{\gamma_r(x)} f(x) = \langle \gamma_r, f \rangle.$$

(So the Fourier transform is simply an inner product with an appropriate character.) The Parseval/-Plancherel identities then have the following form: the first states that

$$\langle f, g \rangle = \langle \widehat{f}, \widehat{g} \rangle_{\ell^2}$$

(so it's a statement that the Fourier transform is a unitary operator), and the second that $\|f\|_2 = \left\|\widehat{f}\right\|_{\ell^2}$.

We'll now prove these identities using this notation.

Proof of Fourier inversion. The set of characters $\{\gamma_r \mid r \in \mathbb{F}_p^n\}$ forms an orthonormal basis of the space of functions $\mathbb{F}_p^n \to \mathbb{C}$ with respect to $\langle -, - \rangle$. To check this, we need to see that

$$\langle \gamma_s, \gamma_r \rangle = \mathbb{E}_x \omega^{(s-r)\cdot x} = \begin{cases} 1 & \text{if } s = r \\ 0 & \text{otherwise.} \end{cases}$$

If $s = r$ then every root is 1; otherwise the sum covers every root of unity equally many times, and is therefore 0. (And it's really a basis because there are as many of them as the dimension of the space.) This is also sometimes called the *Fourier basis*.

Now if you have an orthonormal basis, you can do a change of coordinates — given an arbitrary function $f: \mathbb{F}_p^n \to \mathbb{C}$, imagine we're trying to write f with respect to the Fourier basis. What are its coordinates? Those are the inner products of f with the basis vectors — the coordinate of f at γ_r is equal to $\langle \gamma_r, f \rangle$, which we just saw is $\widehat{f}(r)$.

So if we write f in terms of its basis coordinates, then we get

$$f = \sum_r \widehat{f}(r) \gamma_r.$$

This is precisely the Fourier inversion formula. So the Fourier inversion formula is nothing more than a change of basis (in fact, the Fourier transform is a change of basis). □

Proof of Parseval. We saw above that the Fourier basis is orthonormal, so we can evaluate $\langle f, g \rangle$ in terms of this new basis — we have

$$\langle f, g \rangle = \left\langle \sum_r \widehat{f}(r) \gamma_r, \sum_r \widehat{g}(r) \gamma_r \right\rangle = \sum_r \overline{\widehat{f}(r)} \widehat{g}(r)$$

by orthonormality. (We can also prove this by expanding everything out, but what's happening underneath is the orthonormality of the γ_r's, and this is a more conceptual way to see these identities.) □

Another important operation that comes up is convolution.

Definition 18.9. Given two functions $f, g: \mathbb{F}_p^n \to \mathbb{C}$, we define their *convolution* $f * g: \mathbb{F}_p^n \to \mathbb{C}$ as

$$(f * g)(x) = \mathbb{E}_{y \in \mathbb{F}_p^n} f(y) g(x - y).$$

Another way to view this as

$$(f * g)(x) = \mathbb{E}_{x=y+z} f(y) g(z)$$

(where the expectation is over all pairs (y, z) such that $x = y + z$).

The convolution has several different interpretations.

- If f is supported on A and g is supported on B, then $f*g$ is supported on $A+B = \{a+b \mid a \in A, b \in B\}$. Next chapter, when we study sumsets, this will be quite important.
- If W is a subspace of $\mathbb{F}_p^n$ and we let μ_W be the function

$$\mu_W = \frac{p^n}{|W|}\mathbf{1}_W$$

(where μ_W has average 1 and its mass is uniformly distributed on W), then given a function $f\colon \mathbb{F}_p^n \to \mathbb{C}$, we can consider the convolution $f * \mu_W$. Now g is μ_W, which is supported on this subspace. If we want to evaluate at x, we let z range over W. So this is obtained by averaging f on each W-coset. This is a concept we saw many times in graph theory — in our proofs with Szemerédi's regularity lemma we averaged over parts.

We can think of this as some sort of smoothing — if we start out with an arbitrary, maybe rough, function f and we smooth it with μ_W, it becomes much more well-behaved (in the sense it's constant on every W-coset).

There's a real-analysis version of smoothing — if we take a function that's very rough (it doesn't have a lot of regularity) and we convolve it with something very smooth (like a Gaussian or bump function), then what we get is a smooth version of the first function. What's happening here is an arithmetic analog of this picture.

The relevance of convolutions to 3-APs comes from the convolution identity.

Proposition 18.10 (Convolution identity)

For all $f, g\colon \mathbb{F}_p^n \to \mathbb{C}$ and $r \in \mathbb{F}_p^n$, we have $\widehat{f * g}(r) = \widehat{f}(r)\widehat{g}(r)$.

In other words, the Fourier transform changes convolution to multiplication. The proof is left as an exercise. (It's pretty straightforward — if you write out the expressions and multiply them, you'll see that all the terms that should cancel do cancel.)

§18.2 3-AP densities

Definition 18.11. For three functions $f, g, h\colon \mathbb{F}_p^n \to \mathbb{C}$, let $\Lambda(f,g,h) = \mathbb{E}_{x,y} f(x)g(x+y)h(x+2y)$.

This is a triliear form of the 3-AP density operator.

Definition 18.12. Let $\Lambda_3(f) = \Lambda(f,f,f)$.

If f is an indicator function of some set, this quantity is supposed to capture the number of 3-APs — specifically, for a set A we have

$$\Lambda_3(\mathbf{1}_A) = p^{-2n}\left|\{(x,y) \mid x, x+y, x+2y \in A\}\right|,$$

which we'll refer to as the 3-*AP density* of A. (Here we do allow $y = 0$.)

The following identity relates the Fourier transform with 3-AP counts, and will play a central role in our proofs.

Proposition 18.13

If p is an odd prime, for all $f, g, h\colon \mathbb{F}_p^n \to \mathbb{C}$, we have

$$\Lambda(f,g,h) = \sum_{r \in \mathbb{F}_p^n} \widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r).$$

This relates the Λ operator (a triliear form of the 3-AP operator) to values of the Fourier transform of the individual functions. This semantically looks similar to the convolution identity, and in fact we can prove it using the convolution identity.

Proof. There's a fairly straightforward calculation — we can just expand the left-hand side and cancel all the terms that should be cancelled. The left-hand side is equal to

$$\begin{aligned}\text{LHS} &= \mathbb{E}_{x,y} f(x)g(x+y)h(x+2y) \\ &= \mathbb{E}_{x,y}\left(\sum_{r_1}\widehat{f}(r_1)\omega^{r_1\cdot x}\right)\left(\sum_{r_2}\widehat{g}(r_2)\omega^{r_2\cdot(x+y)}\right)\left(\sum_{r_3}\widehat{h}(r_3)\omega^{r_3\cdot(x+2y)}\right),\end{aligned}$$

using the Fourier inversion formula. If we expand the sum (and exchange the summations and expectations), then we get

$$\text{LHS} = \sum_{r_1,r_2,r_3}\widehat{f}(r_1)\widehat{g}(r_2)\widehat{h}(r_3)\mathbb{E}_x\omega^{x(r_1+r_2+r_3)}\mathbb{E}_y\omega^{y(r_2+2r_3)}.$$

The expectations are both simple exponential sums, so their product is 1 if $r_1+r_2+r_3 = r_2+2r_3 = 0$ and 0 otherwise. We can see that this only happens when (r_1,r_2,r_3) are of the form $(r,-2r,r)$. So that's the only case where the sum doesn't vanish, and that finishes the proof. □

Here's another quick proof of this convolution identity that is also nice to remember, for $\mathbb{F}_3^n$ and with $f = g = h = \mathbf{1}_A$ (which is the setting that we'll actually use); this proof generalizes, but we'll just do this simpler case.

Proof. We can write $\Lambda_3(\mathbf{1}_A) = 3^{-2n}\left|\{(x,y,z)\in A^3 \mid x-2y+z=0\}\right|$. But in $\mathbb{F}_3$ we have $-2=1$, which means we can rewrite this as $x+y+z=0$. Now this expression can be written in terms of convolutions — it's $(\mathbf{1}_A * \mathbf{1}_A * \mathbf{1}_A)(0)$ (if we unpack the definition of the convolution). And we can write this using the Fourier inversion formula as

$$\sum_r \widehat{\mathbf{1}_A * \mathbf{1}_A * \mathbf{1}_A}(r).$$

But the Fourier transform changes convolutions to multiplications, so we get $\sum_r \widehat{\mathbf{1}_A}(r)^3$. □

Remark 18.14. In a way, we've seen this before. This essentially counts the number of closed length-3 walks in the directed graph $G = \text{Cay}(\mathbb{F}_3^n, A)$ (how many ways can we take steps x, y, and z in A and end up back at the origin?). We also know that the number of such walks is $\operatorname{tr} A_G^3$, which we can write as $\sum \lambda_i^3$. And what are the eigenvalues of the Cayley graph? They're precisely the Fourier coefficients — this was a point we made some time ago when discussing the eigenvalues of an abelian Cayley graph. In that way, this identity should feel familiar — it's a relation between counting walks on one hand and taking trace moments on the other hand.

All of this was a quick crash course on the relevant Fourier analysis; we'll now see the proof of Roth's theorem (in the finite field setting) using it.

§18.3 Proof of Roth's theorem

Now we'll prove Roth's theorem for $\mathbb{F}_3^n$, in the following form.

Theorem 18.15 (Roth's theorem in $\mathbb{F}_3^n$)
Eveyr 3-AP-free subset of $\mathbb{F}_3^n$ has size $O(3^n/n)$.

Here's the general strategy (it's a very important one, that Roth came up with).

(1) First, we'll note that every 3-AP-free set has a large Fourier coefficient.

(2) Having a large Fourier coefficient implies that there is a density increment on a hyperplane — there is some hyperplane where by restricting to the hyperplane, we can significantly increase the density of our set.

(3) Then we iterate.

This strategy should be somewhat reminiscent of the one we used to prove Szemerédi's regularity lemma by energy increment — if a graph is not regular, we can find a refinement of our partition so that the energy goes up. Here something thematically similar is happening — if the set is 3-AP-free we can get a density increment on some hyperplane, and this finishes because the density is bounded in $[0,1]$.

To carry out this strategy, we first introduce a useful lemma, which can be viewed as a counting lemma for 3-APs (similar to the triangle counting lemma from last chapter).

Lemma 18.16 (3-AP counting lemma)
For $f\colon \mathbb{F}_p^n \to [0,1]$, we have

$$\left|\Lambda_3(f) - (\mathbb{E}f)^3\right| \le \max_{r\neq 0} \left|\widehat{f}(r)\right| \cdot \|f\|_2^2 .$$

The way to think about this for now is that if we have a function f such that all its nonzero Fourier coefficients are small, this is supposed to be a sign of quasirandomness (as with graphs). And that should imply the 3-AP count (a kind of counting statistic) is close to what you would expect in a random set with the correct density (i.e., $(\mathbb{E}f)^3$).

Proof. We just saw that

$$\Lambda_3(f) = \sum_r \widehat{f}(r)^3 = \widehat{f}(0)^3 + \sum_{r\neq 0} \widehat{f}(r)^3.$$

Here $\widehat{f}(0) = \mathbb{E}f$ is the principal term, and all the other terms are supposed to be small — we have

$$\left|\Lambda_3(f) - (\mathbb{E}f)^3\right| \le \sum_{r\neq 0} \left|\widehat{f}(r)\right|^3 .$$

Now we want to bound this by the maximum of $\widehat{f}$; we'll just take out one factor and replace it by a maximum, to get

$$\left|\Lambda_3(f) - (\mathbb{E}f)^3\right| \le \max_{r\neq 0} \left|\widehat{f}(r)\right| \cdot \sum_{r\neq 0} \left|\widehat{f}(r)\right|^2 .$$

We can enlarge the sum a bit to include 0 as well; then it's equal to $\|f\|_2^2$ by Parseval, which finishes the proof. □

This is similar to the calculation relating C_4 counts and eigenvalues. It's a common mistake to try to replace all three factors by the maximum, but that doesn't work — it's often a good idea to leave a ℓ^2 in, which we can handle quite easily.

§18.3.1 Step 1

We'll now perform the first step, that A being 3-AP-free implies it has a large Fourier coefficient.

Lemma 18.17

Let $A \subseteq \mathbb{F}_3^n$, and let $\alpha = |A| \cdot 3^{-n}$. If A is 3-AP-free and $3^n \geq 2\alpha^{-2}$, then there exists $r \neq 0$ such that $|\widehat{\mathbf{1}_A}(r)| \geq \frac{1}{2}\alpha^2$.

So this states that if we have a 3-AP-free set, as long as n is large enough, we can find a large Fourier coefficient.

Proof. Being 3-AP-free implies that $\Lambda_3(A)$, which counts 3-APs, only couts *trivial* 3-APs (which are just elements of A); so

$$\Lambda_3(A) = \frac{|A|}{3^{2n}} = \frac{\alpha}{3^n},$$

since all 3-APs are trivial. Then the counting lemma says that if all the Fourier coefficients are small then the 3-AP count is close to random; being 3-AP-free is not close to random, so one of the Fourier coefficients must be large. More specifically, we have

$$\alpha^3 - \Lambda_3(\mathbf{1}_A) \leq \max_{r \neq 0} \left|\widehat{\mathbf{1}_A}(r)\right| \cdot \|\mathbf{1}_A\|_2^2 = \left|\widehat{\mathbf{1}_A}(r)\right| \cdot \alpha.$$

The left-hand side is equal to $\alpha^3 - \alpha \cdot 3^{-n}$. We assumed that n is large enough that the left-hand side term is at least $\alpha^3/2$ (i.e., the first term is the main one, and the second is small); then this inequality says that there exists r such that $|\widehat{\mathbf{1}_A}(r)| \geq \alpha^2/2$. □

§18.3.2 Step 2

Step 2 is about what it means to have a large Fourier coefficient — we want to say that it implies a density increment on some hyperplane.

Lemma 18.18

Let $A \subseteq \mathbb{F}_3^n$ and $\alpha = |A| \cdot 3^{-n}$. Suppose that $|\mathbf{1}_A(r)| \geq \delta > 0$ for some $r \neq 0$. Then A has density at least $\alpha + \frac{1}{2}\delta$ when restricted to some hyperplane.

Proof. Let's think about how to interpret Fourier coefficients. We'll do this by revisiting the formula

$$\widehat{\mathbf{1}_A}(r) = \mathbb{E}_x \mathbf{1}_A(x)\omega^{-r \cdot x}.$$

We're working over $\mathbb{F}_3$, so the exponent can only have three possible values — 0, 1, or 2 — and we can partition this expectation according to which value the exponent takes. So we can write this as

$$\widehat{\mathbf{1}_A}(r) = \frac{\alpha_0 + \alpha_1\omega + \alpha_2\omega^2}{3},$$

where α_0, α_1, and α_2 are the densities of A on the three cosets of $r^\perp$. (We have a hyperplane $r^\perp$, which has three different cosets; on each coset we have a different value for $r \cdot x$, and we can separate the expectation into the three cosets.)

If $\alpha_0 = \alpha_1 = \alpha_2$ then the right-hand side is zero — so the only way for the left-hand side to be large is if these three numbers are not too similar. But they also average to α; so if they're not too similar to each other, then one should be small and one should be large.

That's the idea, and then we can do a quick calculation to conclude the result. But we'll do the calculation in a specific way, because it'll be useful in the integer setting. We have

$$\alpha = \frac{\alpha_0 + \alpha_1 + \alpha_2}{3}.$$

So by the triangle inequality, we have

$$\begin{aligned} 3\delta &\le \left|\alpha_0 + \alpha_1\omega + \alpha_2\omega^2\right| \\ &= \left|(\alpha_0 - \alpha) + (\alpha_1 - \alpha)\omega + (\alpha_2 - \alpha)\omega^2\right| \\ &\le |\alpha_0 - \alpha| + |\alpha_1 - \alpha| + |\alpha_2 - \alpha|. \end{aligned}$$

Then one of these terms must be large; but we actually want $\alpha_i - \alpha$ to be positive, not just large in magnitude. We can see this by adding another zero-sum term to write

$$3\delta \le \sum_{j=0}^{2}(|\alpha_j - \alpha| + (\alpha_j - \alpha)).$$

Now each summand is nonnegative — $|x| + x = 2\max(x, 0)$. Having a lower bound on this quantity means that one of the terms must be quite positive, which gives the desired bound — there exists j such that $|\alpha_j - \alpha| + (\alpha_j - \alpha) \ge \delta$, which must mean that $\alpha_j - \alpha \ge \delta/2$. □

We can see here how we're using Fourier coefficients — the formula says that you look at the direction of r and get three hyperplanes perpendicular to r. The Fourier coefficient is in some sense a measure of how much discrepancy we have on these three cosets — if the coefficient is large there must be a lot of discrepancy, so we can move to one of those cosets to get an increment.

§18.3.3 Step 3

Combining steps (1) and (2), we have the following.

Lemma 18.19

Let $A \subseteq \mathbb{F}_3^n$ and $\alpha = |A| \cdot 3^{-n}$. If A is 3-AP-free and $3^n \ge 2\alpha^{-2}$, then A has density at least $\alpha + \alpha^2/4$ when restricted to some hyperplane.

Remark 18.20. Note that our hyperplanes don't have to go through the origin — we're allowed to take cosets.

So as long as we have a 3-AP-free set and n is not too small, we can get a density increment on some subspace. And then we can just restrict A to that hyperplane, and we can repeat — because the restricted A is still going to be 3-AP-free. So we can iterate.

We'll need to keep track of how many times we iterate — the density has to be bounded by 1, and that limits the number of steps we'll take. We can see immediately that we get *something*, but we need to calculate what the something is.

Suppose we start with $A \subseteq \mathbb{F}_3^n$ and $V_0 = \mathbb{F}_3^n$, and $\alpha_0 = \alpha = |A| \cdot 3^{-n}$. If we repeat this lemma, after i rounds we will have restricted A to a codimension-i coset subspace V_i (at each step we drop the dimension by 1). Let α_i be the density of A on V_i, i.e.,

$$\alpha_i = \frac{|A \cap V_i|}{|V_i|}.$$

As long as V_i is not too small — i.e., $2\alpha_i^{-2} \le |V_i| = 3^{n-i}$ — we can apply the lemma to obtain the next step V_{i+1} with

$$\alpha_{i+1} \ge \alpha_i + \frac{\alpha_i^2}{4}.$$

We also know that the density always remains at most 1; so how many steps can we have before we definitely run out of room?

We start with α, and all the densities are at least α; so each step increases the density by at least $\alpha^2/4$. This means there must be at most $4/\alpha^2$ rounds (or else we would exceed 1 in density). What does that give us?

Let the number of rounds be $m \le 4/\alpha^2$. At the time when we stop, we must stop because V_i is too small (if the condition is satisfied then we can always continue); this means

$$3^{n-m} < 2\alpha_m^{-2} \le 2\alpha^{-2}.$$

If we work out what this inequality says, we obtain that $n < m + \log_3(2\alpha^{-2})$. Here m is the dominant term, so this is $O(1/\alpha^2)$; this gives us that $\alpha = O(1/\sqrt{n})$.

This is not quite what we claimed — what this shows us is that A has size $O(3^n/\sqrt{n})$, which is slightly shy of what we claimed from earlier. Now we'll see a small trick that boosts this analysis a bit more to get what we actually claimed (though this is already pretty good).

§18.3.4 A finer analysis

At each step, we see that α_i goes up by some amount. The amount that it goes up by actually increases over time — as α gets bigger, the step size does too. So at each step we actually make more gains; that's not captured in our analysis.

Being exact would be nasty; a common idea in analysis is to do a dyadic analysis, where we think about when things double (and don't worry about any other scale).

At each round, α_i increases by at least $\alpha^2/4$, so it takes at most $\lceil 4/\alpha \rceil$ rounds to double from the start. (Let's not worry about going all the way to the end, and only think about how long it takes to double.)

Once it's doubled, we restart the analysis and think about how many rounds it takes to double again — α has doubled, so it now takes $\lceil 2/\alpha \rceil$ rounds to double again. And so on — the next time it takes $\lceil 1/\alpha \rceil$ rounds to double, and then $\lceil 1/2\alpha \rceil$, and so on.

Since the initial density is α, we can double at most $\log_2 1/\alpha$ times (since the density cannot exceed 1). So after we've doubled so many times, the process must stop; the number of rounds from all the doubling is at most

$$\sum_{j \le \log_2 1/\alpha} \left\lceil \frac{2^{2-j}}{\alpha} \right\rceil = O(1/\alpha).$$

If the process terminates after m steps with density α_m, then we can do the same analysis as before to see that

$$|V_m| = 3^{n-m} < \alpha_m^{-2} \le \alpha^{-2}.$$

This gives $n \le m + O(\log \frac{1}{\alpha}) = O(1/\alpha)$; this gains the extra square root factor.

§18.4 Concluding remarks

This finishes the proof of Roth's theorem in $\mathbb{F}_3^n$. We'll now address some questions.

First, how general is this proof — does it work for 4-APs? The short answer is, not quite. There'll be a homework problem explaining why this method doesn't work — in the homework problem we'll see a set that has small Fourier coefficients, but its number of 4-APs is not random-like, so that the 3-AP counting lemma is false for 4-APs. This is a fundamental obstruction — one active area of additive combinatorics is to understand how to deal with 4-APs. The short answer is that you have to move beyond Fourier analysis

to quadratic Fourier analysis; we won't have time to get into it, but there are homework problems hinting towards its direction.

Another comment, which is good to think about: this is a beautiful proof, but there's lots of subtleties. Where can this proof go wrong if we change the statement?

For example, what if instead of considering 3-APs (where $x + y = 2z$), what if we consider $x + y = z$ — what's the largest subset of $\mathbb{F}_3^n$ avoiding solutions to this set?

Well, the answer to that question is not small — we can take all points with first coordinate 1 (a hyperplane not through the origin), which is a pretty large set. So the result is not true if we replace this pattern with 3-APs.

Where does the proof fail? The 3-AP counting lemma remains true. But for 3-AP-free, when we passed to a coset, we could pick a new origin and the set remained 3-AP-free. For this pattern, we can't do that. For example, imagine the case where everything is on one hyperplane (the $\frac{1}{3}$ density example). Then in the first round we'll zone in on that hyperplane; and when we pick a new origin, our set is no longer free of this pattern.

So this pattern is not translation-invariant — the choice of origin matters — while 3-APs are, which allowed us to go to an affine subspace and choose a new origin.

Remark 18.21. What if we looked at $x + y = z + w$? There the right bound is the square root of the ambient space; this can be proven using Cauchy–Schwarz or the pigeonhole principle. But here's an open problem — what's the largest subset of $[N]$ avoiding the pattern $x + 3y = 2z + 2w$? This is very much an open problem. On one hand, the Roth proof works — this is translation-invariant. But the Behrend construction doesn't work at all; and the best lower-bound construction is around $\sqrt{N}$ (whereas the best upper-bound construction is just below N). So there's a big gap between these two bounds.

§19 November 15, 2023

Today we'll prove Roth's theorem in $\mathbb{Z}$.

Theorem 19.1 (Roth)

Every 3-AP-free subset of $[N]$ has size $O(N/\log\log N)$.

This is what Roth originally proved in his paper, and the proof we'll see is essentially Roth's proof. Last time, we proved that every 3-AP-free subset of $\mathbb{F}_3^n$ has size $O(3^n/n)$. The proof today will essentially be an elaboration of the strategy from last time, but adapted to the setting of $\mathbb{Z}$; we'll follow many familiar ideas, but now they'll happen in $\mathbb{Z}$.

§19.1 Fourier analysis in the integers

To get started, we'll need to define notions of Fourier analysis in the integers.

Last time, we defined the Fourier transform of a function defined on a finite field vector space (which can be extended to any finite abelian group).

Definition 19.2. Given a finitely supported function $f\colon \mathbb{Z} \to \mathbb{C}$, we define $\widehat{f}\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ as

$$\widehat{f}(\theta) = \sum_{x \in \mathbb{Z}} f(x) e(-x\theta),$$

where $e(t) = e^{2\pi i t}$.

Finitely supported means that the function takes nonzero values at only finitely many points (so that when we deal with sums, we don't have to worry about convergence). We can equivalently think of $\widehat{f}$ as a 1-periodic function on $\mathbb{Z}$.

You may have first seen Fourier analysis as starting with a 1-periodic function and extracting its Fourier series. Here, that corresponds to the Fourier inversion formula:

Proposition 19.3
We have $f(x) = \int_0^1 \widehat{f}(\theta) e(x\theta)\, d\theta$.

This states that given the Fourier transform, you can recover the original function. (This is the formula you usually see when first seeing this in the context of Fourier series.)

We'll now state the standard facts about this operation. The proofs are essentially the same as last time, maybe with more analysis. (It's easier in the finite field setting because you don't have to worry about convergence issues.)

Theorem 19.4 (Parseval/Plancherel)
Given finitely supported $f, g\colon \mathbb{Z} \to \mathbb{C}$, we have

$$\sum_{x \in \mathbb{Z}} \overline{f(x)} g(x) = \int_0^1 \overline{\widehat{f}(\theta)} \widehat{g}(\theta)\, d\theta.$$

In particular, $\sum_{x \in \mathbb{Z}} |f(x)|^2 = \int_0^1 |\widehat{f}(\theta)|^2\, d\theta$.

Again, this states that the inner product in physical space equals the inner product in the frequency space, if we use the correct normalizations.

Definition 19.5. Given $f, g\colon \mathbb{Z} \to \mathbb{C}$, we define their *convolution* $f * g\colon \mathbb{Z} \to \mathbb{C}$ by

$$(f * g)(x) = \sum_{y \in \mathbb{Z}} f(y) g(x - y).$$

Similarly to what we saw last time, the convolution plays well with the Fourier transform.

Proposition 19.6
We have $\widehat{f * g}(\theta) = \widehat{f}(\theta)\widehat{g}(\theta)$.

So in other words, $\widehat{f * g} = \widehat{f}\widehat{g}$ — the Fourier transform turns convolution into multiplication.

All of this is fairly straightforward to check; we won't do it here.

We also saw an identity, somewhat related to the convolution identity, which is relevant to counting 3-APs.

Definition 19.7. For $f, g, h\colon \mathbb{Z} \to \mathbb{C}$, we define $\Lambda(f, g, h) = \sum_{x,y} f(x) g(x + y) h(x + 2y)$.

This is the operator that corresponds to counting 3-term arithmetic progressions. In the case that's relevant to us, where all three functions are the same, we write $\Lambda_3(f) = \Lambda(f, f, f)$.

This function plays well with the Fourier transform.

Proposition 19.8
We have $\Lambda(f, g, h) = \int_0^1 \widehat{f}(\theta)\widehat{g}(-2\theta)\widehat{h}(\theta)\, d\theta$.

We saw a completely analogous formula last time, in the context of finite field vector spaces; if you use the appropriate measure and the right translation, then you also have this formula here (we're not going to prove it, but you can use the same proof — you can expand the formula for each of these terms on the right-hand side using the definition, and expand and collect the terms that don't cancel).

§19.2 The strategy

Our goal for today is to prove Roth's theorem. Let's recall the strategy we used last time to prove Roth's theorem in $\mathbb{F}_3^n$:

(1) First we showed that if we have a set that's not too small and is 3-AP-free, then it must have some large Fourier coefficient.

(2) Next, we showed that if you start with a large Fourier coefficient, then we can get a density increment. Last time, the density increment was from restricting to a hyperplane (we can find some hyperplane where the density goes up substantially). Now we're working in $\mathbb{Z}$, and there's no hyperplanes anymore. So instead what we'll do is restrict onto some *subprogression* — we start with a progression $\{1, \dots, N\}$, and we restrict to some (still fairly long) arithmetic progression inside the original set, such that the density goes up substantially.

(3) We iterate; this can only go on for some number of times, and we see what happens when the iteration necessarily has to stop.

§19.3 A counting lemma

Last time, one of the things we needed was a counting lemma for 3-APs; here's what such a counting lemma looks like.

Lemma 19.9
For (finitely supported) $f, g \colon \mathbb{Z} \to \mathbb{C}$, we have

$$|\Lambda_3(f) - \Lambda_3(g)| \le 3 \left\| \widehat{f - g} \right\|_\infty \max\{\|f\|_{\ell^2}^2, \|g\|_{\ell^2}^2\}.$$

So f and g differ by a quantity that's supposed to be small if the Fourier transforms of f and g are close together. To see what this notation means, $\left\|\hat{f}\right\|_\infty = \sup_\theta |\widehat{f}(\theta)|$, and $\|f\|_{\ell^p} = (\sum_{x \in \mathbb{Z}} |f(x)|^p)^{1/p}$. (The notation ℓ^p, as opposed to L^p, denotes a sum.)

In what ways is this different from what we had last time? Now we're comparing two functions to each other, instead of comparing one function to random — last time g was just a constant α (which was supposed to be the density of f). Now we allow more flexible g; we'll see that this will come up.

How do we interpret this quantity on the right-hand side? Last time, when g was a constant function, the Fourier transform of a constant function is concentrated at 0 (and is 0 everywhere else). Last time we had a quantity of the form $\max_{r \neq 0} |\widehat{g}(r)|$. The point is that this is the same as $\max_r |\widehat{(g - \mathbb{E}g)}(r)|$ (where the

maximum is now over all r, including 0) — because subtracting the expectation knocks out the possibility of r being 0 (in which case the expression in the expectation is 0). That's what this 'de-meaning' does.

Proof. The proof is reminiscent of the proof of the triangle counting lemma for graphons — we're going to imagine changing from $\Lambda(f,f,f)$ to $\Lambda(g,g,g)$ through three steps, where at each step we change one of the inputs — we can write

$$\Lambda(f,f,f) - \Lambda(g,g,g) = \Lambda(f-g,f,f) + \Lambda(g,f-g,f) + \Lambda(g,g,f-g).$$

We'll bound each of these separately (the calculation is identical). For the first, we have

$$|\Lambda(f-g,f,f)| = \left| \int_0^1 \widehat{f-g}(\theta) \widehat{f}(-2\theta) \widehat{f}(\theta)\, d\theta \right|.$$

Now we can take out the $\widehat{f-g}$ term while keeping the others present — this is at most

$$\left\| \widehat{f-g} \right\|_\infty \int |\widehat{f}(-2\theta)||\widehat{f}(\theta)|\, d\theta.$$

Now by Cauchy–Schwarz, we can decompose this as

$$\le \left\| \widehat{f-g} \right\|_\infty \left(\int_0^1 |\widehat{f}(-2\theta)|^2\, d\theta \right)^{1/2} \left(\int_0^1 |\widehat{f}(\theta)|^2\, d\theta \right)^{1/2},$$

and these two integrals are $\|f\|_{\ell^2}$ by Plancherel. (You can do the same for the other two terms; the worst case corresponds to taking the maximum.) □

§19.4 Proof of Roth

Last time, we worked with $\mathbf{1}_A$, and looked at $\widehat{\mathbf{1}_A}(r)$ for some nonzero r. Here we can try looking at $\widehat{\mathbf{1}_A}(\theta)$ for $\theta \neq 0$, but that isn't great — this is a continuous function of θ, so excluding $\theta = 0$ doesn't help you that much. In particular, $\widehat{\mathbf{1}_A}(0) = |A|$ is a pretty big value; so we'll have $\widehat{\mathbf{1}_A}(\theta) \approx |A|$ for $\theta \approx 0$. This is not so great — it gives us no information.

We can try excluding a small interval around 0, but that's difficult to work with. A better approach is to de-mean the indicator function by a constant (corresponding to the mean); that will in effect get rid of the contributions from θ close to 0.

So we will look at the de-meaned indicator function

$$\mathbf{1}_A - \alpha \mathbf{1}_{[N]},$$

where $\alpha = |A|/N$. The intuition is that when A is close to random-like, this function is suppossed to be very uniform, in the sense of having small Fourier coefficients (and that is true whether or not θ is close to 0).

§19.5 The first step

We'll now prove a sequence of lemmas corresponding to the strategy, analogous to what we did last time.

The first is that having a 3-AP-free set implies having a large Fourier coefficient.

Lemma 19.10

Let $A \subseteq [N]$ be 3-AP-free, and let $|A| = \alpha N$. If $N \geq 5\alpha^{-2}$, then there exists θ such that

$$\left| \sum_{x=1}^{N} (\mathbf{1}_A - \alpha \mathbf{1}_{[N]})(x) e(\theta x) \right| \geq \frac{\alpha^2}{10} N.$$

(This quantity is just the Fourier transform.)

We need N to not be too small, or else the proof won't work; this will matter at the end when we figure out how small the density can be. We shouldn't worry about the random constants like 10.

So this states that being 3-AP-free means there exists some θ such that the de-meaned indicator function has a large Fourier coefficient at θ.

Proof. Since A is 3-AP-free, we know that $\Lambda_3(\mathbf{1}_A) = |A|$ — there are no 3-APs, so the only nontrivial contributions in the definition are when $y = 0$ (corresponding to the trivial 3-APs), of which there are exactly $|A|$.

On the other hand, if A were close to random (so all these Fourier coefficients were small), then the 3-AP counting lemma would tell us that A has lots of 3-APs (close to the random number); these would conflict, and that's where the lemma morally arises.

On the other hand, it's a combinatorial exercise to estimate $\Lambda_3(\mathbf{1}_{[N]})$; you can imagine picking the first term and the last term with the same parity. This gives

$$\Lambda_3(\mathbf{1}_{[N]}) = \left\lfloor \frac{N}{2} \right\rfloor^2 + \left\lceil \frac{N}{2} \right\rceil^2 \geq \frac{N^2}{2}.$$

Now we apply the counting lemma; this gives

$$\Lambda_3(\alpha \mathbf{1}_{[N]}) - \Lambda_3(\mathbf{1}_A) \leq 3\alpha N \left\| \widehat{\mathbf{1}_A - \alpha \mathbf{1}_{[N]}} \right\|_\infty.$$

On the other hand, by the calculation we just did,

$$\Lambda_3(\alpha \mathbf{1}_{[N]}) - \Lambda_3(\mathbf{1}_A) \geq \frac{\alpha^3 N^2}{2} - \alpha N.$$

Using the fact that $N \geq 5\alpha^{-2}$ is somewhat large, we see that the first term dominates — dividing out by $3\alpha N$, we have

$$\frac{1}{3\alpha N} \left(\frac{\alpha^3 N^2}{2} - \alpha N \right) \geq \frac{1}{6}\alpha^2 N - \frac{1}{3} \geq \frac{1}{10}\alpha^2 N,$$

which finishes the proof. □

There's some calculation here, but the moral is the same as before — if this were not true, it would suggest that A is random-like, which would tell us that A has close to the random number of 3-APs; that's not true because A is 3-AP-free.

§19.6 The second step

Now we have a large Fourier coefficient; how should we think about the second step?

Let's recall what happened last time, with $A \subseteq \mathbb{F}_3^n$ — suppose that $|\widehat{\mathbf{1}_A}(r)|$ was large. Then looking in the direction of r, we could partition $\mathbb{F}_3^n$ into the three cosets of $r^\perp$. This Fourier coefficient being large

implies that the densities of A on these three slices are not too similar to each other — there has to be some discrepancies, which corresponds to one of these having many more elements of A than average.

Now we're in a much more continuous setting; how can we interpret what it means for this function to be large?

In the $\mathbb{F}_3^n$ setting, it was helpful that there were only 3 cosets, so we could say one of them was large. Here we want to do something similar and discretize the problem.

To make our life easier, let's pretend for now that θ is a rational number with a small denominator b, and let's think about the character $x \mapsto e(x\theta)$. How does this behave? This function is much like the character we saw earlier — it's periodic mod b. So if b is fixed (e.g. 3), then it's the same situation as before — we've partitioned our set into b sets, and there has to be some discrepancy between these parts.

So if θ is rational with small denominator, we're in a completely analogous situation to earlier. But b could be very large, or θ could be irrational. But that's okay — everything behaves rather continuously, so as long as we can *approximate* θ by a number with a small denominator, we're good to go. So our goal now is to first approximate θ by a rational number $\frac{a}{b}$ with small denominator, and then see what happens — similar to last time — and then show that this approximation doesn't change much.

We use $\|\theta\|_{\mathbb{R}/\mathbb{Z}}$ to mean the distance from θ to the nearest integer. There's a useful but simple lemma:

Lemma 19.11 (Dirichlet's lemma)
For every $\theta \in \mathbb{R}$ and $0 < \delta < 1$, there exists a positive integer $d \le \delta^{-1}$ such that $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \le \delta$.

(This is getting towards what we want — θ is pretty close to a rational number with a small denominator.)

Proof. We use the pigeonhole principle — let $m = \lfloor 1/\delta \rfloor$. Among the numbers 0, θ, 2θ, $\dots$, $m\theta$, we can think of them as lying on a circle (by taking their fractional parts). If we have $m+1$ numbers on the circle, then two must be very close to each other — by pigeonhole there exist $i < j$ such that the fractional parts of $i\theta$ and $j\theta$ differ by at most δ, and then we can set $d = |i - j|$. □

Sometimes this is interpreted in the following way: if we start with some step size at 0 and take a walk along the circle, after not too many steps we should get back close to 0.

In the finite field setting we were able to exactly partition the space into three cosets, each of which was a hyperplane. We want to do something similar in the integer setting — some kind of partition into sets where this character has (roughly) the same value. So informally, given θ, we want to partition $[N]$ into subprogressions (i.e., we want each of our sets to be an AP) such that on each, the function $x \mapsto e(x\theta)$ is roughly constant. (In the finite field case, we could partition into three subspaces where the character was exactly constant; here we just want 'roughly' constant.)

Lemma 19.12
Let $0 < \eta < 1$ and $\theta \in \mathbb{R}$. Suppose that $N \ge (4\pi/\eta)^6$. Then one can partition $[N]$ into subprogressions P_i, each with length $N^{1/3} \le |P_i| \le N^{1/3}$, such that for each i, we have

$$\sup_{x,y \in P_i} |e(x\theta) - e(y\theta)| < \eta.$$

We have some parameters and constants; we should ignore the arbitrary-looking constants for the most part. For example, η is a parameter we'll adjust later; we should ignore 4π and 6. This states that we can partition $[N]$ into relatively long progressions such that on each, $x \mapsto e(x\theta)$ is roughly constant (up to some small tolerance, given by η).

Proof. By the Dirichlet lemma, there exists $d < \sqrt{N}$ such that $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \le 1/\sqrt{N}$. We're going to partition $[N]$ greedily into progressions with common difference d, of length between $N^{1/3}$ and $2N^{1/3}$.

The idea is we want to partition $[N]$ into progressions on which $x \mapsto e(x\theta)$ is roughly constant. It's not too hard to see how you can make this roughly constant on progressions — on a progression x and y differ by some number of steps of size d, and we want to choose d such that $e(d\theta)$ is pretty close to 1. So that's what we do — we pick d such that θd is close to an integer, and divide our set into sets of common difference d.

Once you pick d, you can first look at what happens mod d; in each residue class you'll get a progression of length N/d. This may be too long; then you can chop it up into progressions of the appropriate size.

The rest is a calculation — we have

$$|e(x\theta) - e(y\theta)| \le |P_i|\,|e(d\theta) - 1|$$

(the length of the progression times the discrepancy at each step). We have $|P_i| \le 2N^{1/3}$, and $|e(d\theta) - 1| \le 2\pi/\sqrt{N}$ (by the fact that $d\theta$ is close to an integer). This gives $|e(x\theta) - e(y\theta)| \le \eta$, as desired. □

So this says we can partition our big progression $[N]$ into a small number of progressions, each of which is pretty long, such that the character we got from the previous step is roughly constant on each.

Remark 19.13. The proof is more technical than what we saw last time, but the broad strategy is the same. This reinforces something we said last time — the finite field model is often a great playground to try ideas, because the execution is often simpler while many of the ideas remain similar.

The next step will tell us that 3-AP-free implies a density increment.

Lemma 19.14

Let $A \subseteq [N]$ be a 3-AP-free set with $|A| = \alpha N$, with $N \ge (16/\alpha)^{12}$. Then there exists a subprogression $P \subseteq [N]$ with $|P| \ge N^{1/3}$ and such that

$$\frac{|A \cap P|}{|P|} \ge \alpha + \frac{\alpha^2}{40}.$$

So the density of A was originally α, and on this progression, it goes up by something on the order of α^2. This is the density increment we're looking for, and once we have this we can iterate to get the density higher and higher.

Proof. There exists θ such that

$$\left|\sum_{x=1}^{N} (\mathbf{1}_A - \alpha)(x) e(x\theta)\right| \ge \frac{\alpha^2}{10} N.$$

Now we apply the previous lemma with $\eta = \alpha^2/20$; we need to check that the hypothesis to the lemma (of N being reasonably large) is satisfied, and that is indeed true. This gives a partition of $[N]$ into progressions $P_1, \ldots, P_k$ such that each progression has size roughly $N^{1/3}$ (more precisely $N^{1/3} \le |P_i| \le N^{2/3}$), and furthermore, on each progression we have $|e(x\theta) - e(y\theta)| \le \alpha^2/20$.

Now we're going to combine these two facts — on each progression the character doesn't vary too much, so we can pretend it's constant; then we can decompose into a number of sets, where one should have a density increment.

On each progression P_i, we see that

$$\left|\sum_{x \in P_i} (\mathbf{1}_A - \alpha)(x) e(x\theta)\right| \le \left|\sum_{x \in P_i} (\mathbf{1}_A - \alpha)(x)\right| + \frac{\alpha^2}{20}\,|P_i|$$

(the first term is what we get by pretending $e(x\theta)$ is constant, and the second comes from keeping track of the error term). This means

$$\frac{\alpha^2}{10}N \le \left|\sum_{x=1}^{N}(\mathbf{1}_A-\alpha)(x)e(x\theta)\right| \le \sum_{i=1}^{k}\left|\sum_{x\in P_i}(\mathbf{1}_A-\alpha)(x)e(x\theta)\right|$$

(decomposing into progressions). On each progression $e(x\theta)$ is roughly constant, so we can take it out and add in the error estimate $\frac{\alpha^2}{20}|P_i|$; then we can collect these together and get that

$$\frac{\alpha^2}{10}N \le \sum_{i=1}^{k}\left|\sum_{x\in P_i}(\mathbf{1}_A-\alpha)(x)\right| + \frac{\alpha^2 N}{20}.$$

The first term must be quite large. But these things are basically the number of elements A has in each progression, and rearranging, we find that

$$\frac{\alpha^2}{20}\sum_{i=1}^{k}|P_i| \le \sum_{i=1}^{k}||A\cap P_i|-\alpha|P_i||$$

(where $N=\sum|P_i|$, and $|A\cap P_i|-\alpha|P_i|$ is the difference between the size of A on P_i and its expectation).

We want to show that if we have lots of fluctuations, then one of these is large. We can immediately conclude that one of them must be large in *absolute value*, but we actually want to show it's large. But this is not hard to show because the insides of the absolute values need to sum to 0, which means they need to balance each other out.

So we can use the same trick last time — we can write the right-hand side as

$$\sum_{i=1}^{k}||A_i\cap P_i|-\alpha|P_i|| + |A_i\cap P_i|-\alpha|P_i|$$

(here we're writing $x+|x|$ — the x's average out to 0, so we can throw them in). Then this gives that there exists i such that

$$\frac{\alpha^2}{20}|P_i| \le ||A\cap P_i|-\alpha|P_i|| + (|A\cap P_i|-\alpha|P_i|).$$

The right-hand side is always nonnegative, so we can conclude that

$$\frac{\alpha^2}{40}|P_i| \le |A\cap P_i|-\alpha|P_i|,$$

which finishes the proof of the lemma. □

§19.7 Iteration

So we've found a subprogression that is still pretty long (around the cube root of the original size) on which the density of A increases substantially. Now we're just going to repeat that procedure — like last time, we zone in on the subprogression (which for our purposes looks like $[N']$ for some smaller N'). Then A is still 3-AP-free, so we can keep iterating, until we get stuck.

This certainly has to stop after some number of steps (the density can't increase forever); and when we stop, it must be because N is too small (violating the condition in our lemma).

So we start with $\alpha_0=\alpha$ and $N_0=N$. After i iterations, we arrive at a progression of length N_i where

$$\frac{|A\cap P_i|}{|P_i|}=\alpha_i.$$

At each step the length of the progression goes down, but not too much — we have $N_{i+1} \geq N_i^{1/3}$ — whereas the density increases substantially — we have $\alpha_{i+1} \geq \alpha_i + \alpha_i^2/40$.

Then counting the number of steps is the same as last time (we could do this naively bounding α_i by α, but we get a better analysis by being more careful) — we double α_i from α_0 after at most $\lceil 40/\alpha \rceil$ iterations, and then we double again after $\lceil 20/\alpha \rceil$ iterations, and then again after at most $\lceil 10/\alpha \rceil$ iterations, and so on. We can only double at most $\log_2 1/\alpha$ times (since the density is bounded by 1). So the total number of iterations is at most

$$m \leq \sum_{i=1}^{\log(1/\alpha)} \left\lceil \frac{40}{2^i \alpha} \right\rceil = O(1/\alpha).$$

(The main term is $O(1/\alpha)$; we get some rounding from the ceilings, but that contributes very little since there are very few steps.)

The process only terminates when the size condition on N is violated — i.e., when $N^{1/3^m} \leq N_m < (16/\alpha_m)^{12} \leq (16/\alpha)^{12}$. We can rearrange everything to see that $N \leq (16/\alpha)^{e^{O(1/\alpha)}}$. The $16/\alpha$ no longer matters (it's the exponential term that's dominant), and we get that

$$\frac{|A|}{N} = \alpha = O\left(\frac{N}{\log\log N}\right)$$

(because the bound on N is double-exponential in $1/\alpha$).

§19.8 Concluding remarks

This concludes our proof of Roth's theorem. This is one of the foundational results in modern additive combinatorics. This is in fact the second proof we saw in the class — the first was by Szemerédi's regularity lemma, and now we've seen another proof using Fourier analysis.

Remark 19.15. Is there a reason why this proof gets us much better quantititive bounds compared to the regularity proof? The regularity lemma itself has very poor bounds, but what we used is the graph removal lemma; we only know how to prove the graph removal lemma using the regularity method. You could say that there you're looking at the whole graph; whereas here you don't care about structure outside the sets you're zooming in on (which is much more efficient).

But also, we don't actually know whether you *need* tower-type bounds for the graph removal lemma — the regularity lemma does need tower-type bounds, but for the removal lemma this is still open.

We mentioned at the beginning of the class that the $O(N/\log\log N)$ bound has been improved many times. This year there was a recent breakthrough — a bound so good it's almost like the Behrend construction. The best result (from earlier this year) is $O(\frac{N}{e^{(\log N)^c}})$. This is a huge surprise — a lot of previous work had been done, but was very far from the Behrend bound.

What happens if we compare the two settings (of $\mathbb{Z}$ vs. the finite field setting)? Are the bounds we proved parallel? Today's bound is worse — we lost an extra log. Let's think about why — if $3^N = N$, then the bound we proved last time is $N/\log N$. So there is a difference.

This difference comes from the fact that last time we passed down to a hyperplane, so the size of the space went down by a factor of 3 (which is a constant). Today the size at each step went down by a cube root; that's a much bigger loss. This is significant — it explains the difference between $N/\log N$ and $N/\log\log N$.

For a very long time, there was an effort to try to mimic the $N/\log N$ type of result in the setting of integers. This was successfully done, and there were different ideas; here's one important idea, introduced by Bourgain, which concerns *Bohr sets*. (This is not named after Nils Bohr, but his brother, who was a mathematician studying almost-periodic functions.)

Today we passed down from progressions to subprogressions; last time we passed down from spaces to subspaces. These seem similar, but they are quantitatively different. It turns out there's a way to go from progressions to something more analogous to the subspace version.

The idea is that in $\mathbb{F}_3^n$, we passed down from the entire set to a hyperplane, which is defined by $\{x \mid \langle x, r\rangle = c\}$ for some r. The idea in the world of progressions is to do something similar.

In the integers, we pass from the original space to a set

$$\{x \in [N] \mid \|x\theta_j\|_{\mathbb{R}/\mathbb{Z}} \le \varepsilon \text{ for all } j = 1, \dots, k\}.$$

This is analogous to a subspace, which is a set $\{x \in \mathbb{F}_3^n \mid x \cdot r_j = 0 \text{ for all } j = 1, \dots, t\}$. This is an analogous concept, called a *Bohr set*; it turns out that quantitatively, it's better to use these kinds of objects compared to progressions. (Progressions are nicer combinatorially but analytically don't form a perfect analogy; these sets are better quantitatively to work with.)

It's still not easy to translate what we saw last time to Bohr sets; a lot of work was done, and the bounds went from $N/(\log N)^c$ for some small c, to the power getting bigger and bigger; a few years ago they finally proved $N/(\log N)^{1+\varepsilon}$, which was a huge breakthrough.

So today we used progressions, as Roth did; this is simpler and combinatorially easier to explain. But to do better you have to think about these other objects which are more complicated, but in some sense more natural and a better analogy. (We'll see Bohr sets again in the next chapter.)

Next time, we'll begin with another proof of Roth's theorem in finite fields, but it'll produce an *exponentially* better bound — this will use the polynomial method. This was a huge breakthrough when it came out (in 2017). However, that proof is very specific to the finite field setting. (It is not known whether any of these ideas can be at all translated to the integer setting.)

§20 November 20, 2023

Today we'll look at another proof of Roth's theorem in the finite field setting, using the *polynomial method*. A couple of lectures ago, we saw a proof of Roth's theorem in finite fields using Fourier analysis. For a long time this was the approach everyone used, and there were a few small (but important) improvements over the bound we saw.

There was a huge breakthrough in around 2017, when Croot–Lev–Pach found a method that completely transformed the subject. They solved the problem for $(\mathbb{Z}/4\mathbb{Z})^n$, which is not quite the finite field setting (since $\mathbb{Z}/4\mathbb{Z}$ is not a field); still, they got an exponential improvement, which was very exciting. In less than a week, Ellenberg–Gijswijt found that using a variant of these ideas, we can get $\mathbb{F}_p^n$ for any fixed prime p.

Theorem 20.1

Every 3-AP-free subset in $\mathbb{F}_3^n$ has size at most 2.76^n.

Notably, this is exponentially smaller than the trivial bound of 3^n. This theorem was a shock to the community — for a long time everyone struggled pushing the bounds bit by bit, and it was not clear whether you could get an exponential bound. For contrast, in the integers there's the Behrend bound; but in finite fields the Behrend construction doesn't work. The best construction comes from taking some specific example (which is still fairly large) and tensoring it up; that gives roughly 2.21^n. So before this result, it wasn't clear whether you could get a base less than 3; it was quite a surprise that you can.

We'll follow a presentation given by Tao on his blog (where he explains the ideas of these papers in a way that's quite transparent).

§20.1 Slice rank

We'll consider the notion of *rank*. In linear algebra, you can view a matrix as a function $F: A \times A \to \mathbb{F}$ (where A is a finite set). We say a matrix has rank 1 if we can decompose this function as $F(x, y) = f(x)g(y)$. In general, the *rank* of a matrix is the minimum k such that F is the sum of k rank-1 matrices.

So that's the classic linear algebraic definition of rank. But what if you have more than 2 coordinates in your function — what about a function $A \times A \times A \to \mathbb{F}$? It's enough to say what are all the rank 1 functions; then the rank is the number of rank 1 functions we need to write F as a sum.

One common notion is the *tensor rank* — F has rank 1 if $F(x, y, yz) = f(x)g(y)h(z)$. This is important in many other contexts, but it's not the one we'll discuss.

For this application, we'll consider a slightly different notion of rank, known as *slice rank*.

Definition 20.2. We say a function $F: A \times A \times A \to \mathbb{F}$ has *slice rank* 1 if it can be written as $F(x, y, z) = f(x)g(y, z)$ (for all x, y, and z), or $F(x, y, z) = f(y)g(x, z)$, or $F(x, y, z) = f(z)g(x, y)$.

So F has slice rank 1 if it can be written as a product of two functions, where one only depends on one variable and the other only depends on the other two.

Definition 20.3. The *slice rank* of a function $F: A \times A \times A \to \mathbb{F}$ is the minimal k such that F can be written as a sum of k slice rank 1 functions.

Let's play around with some basic properties of slice rank. First, we have a trivial upper bound:

Lemma 20.4 (1)

Every function $F: A \times A \times A \to \mathbb{F}$ has slice rank at most $|A|$.

Proof. Pictorially, think of the function as a 3-dimensional 'matrix' — we have a cube, and each entry of the cube is a value of F. We want to write F in terms of various 'slices,' where each slice F_a is the restriction of F to the slice where $x = a$, and 0 elsewhere. Each F_a has slice rank 1, since we can decompose it as $F_a(x, y, z) = \mathbf{1}_a(x)F(a, y, z)$ (where the first only depends on x, and the second on y and z). So we can decompose $F = \sum F_a$, as desired. □

The next fact is an elementary fact from linear algebra about when you can find vectors with large support in a subspace.

Lemma 20.5 (2)

Every k-dimensional subspace of $\mathbb{F}^n$ contains a point with at least k nonzero coordinates.

Another way to think about this is that if I give you a $k \times n$ matrix, and the rows are linearly independent, then I can find some linear combination of the rows such that the resulting vector has support size at least k. Some linear combinations might have too small support, but I can always find *some* with large support.

This should intuitively be clear, but you have to work a little to prove it.

Proof. Let our $k \times n$ matrix be M, so that $\mathrm{Span}(M)$ is the k-dimensional subspace we're given. We know $\mathrm{rank}(M) = k$; this means it has some invertible $k \times k$ submatrix (where by a 'submatrix' we mean that we can take k different columns, and look at the matrix generated by those k columns). This can be seen by doing Gaussian elimination on the rows — eventually you get some pivot entries, and they form an invertible $k \times k$ matrix.

Then we can look at this $k \times k$ submatrix, and we can generate 1's on these k coordinates (which certainly makes them nonzero) — since it's invertible, it spans all possible entries on these k coordinates. □

The next fact considers the slice rank of a diagonal tensor.

Lemma 20.6 (3)

If $F: A \times A \times A$ has the property that $F(x, y, z) \neq 0$ if and only if $x = y = z$, then the slice rank of F is $|A|$.

A diagonal matrix has 0's off the diagonal; if all its diagonal entries are nonzero, then it has full rank. This is a corresponding statement.

Proof. We already know that the slice rank is *at most* A, so it suffices to prove the lower bound — that we cannot write F as a sum of *fewer* than $|A|$ rank 1 functions.

Suppose that there exists a way to write F as a sum of functions of the forms $f(x)g(y, z)$, $f(y)g(x, z)$, and $f(z)g(x, y)$ (the elementary rank 1 functions). We have a number of summands; suppose that we have m_1 of the first type, m_2 of the second, and m_3 of the third.

Then by Lemma 2, there exists a function $h: A \to \mathbb{F}$ which is orthogonal to all the f's from the third type (where by *orthogonal* we mean that $\sum f(x)h(x) = 0$) and has large support — specifically, $|\operatorname{supp} h| \geq |A| - m_3$. (Given a bunch of functions, we can look at their orthogonal complement; that's a subspace of dimension $|A| - m_3$, and Lemma 2 says we can find an element whose support is at least the rank of this subspace.)

Now let's multiply h against the function we're given and sum over z; let

$$G(x, y) = \sum_{z \in A} F(x, y, z)h(z).$$

What happens? Every contribution of the *third* type gets annihilated, so only the first two types of summands remain. (The third type of summand gets annihilated because we chose h to be orthogonal to all the f's from the third type.)

The first type gets turned into a function taking (x, y) as input and outputting $\sum_z f(x)g(y, z)h(z)$. We can collect the sum over z and write this as $f(x)\widetilde{g}(y)$ for some new function $\widetilde{g}$; in particular, this is a rank 1 function (in the linear algebraic sense, since we only have two variables).

The same calculation works for the second type.

So $\operatorname{rank} G \leq m_1 + m_2$ (where G is a function $A \times A \to \mathbb{F}$, so it's a traditional matrix).

But what was G? Well, F was a diagonal function; so we see that G is also a diagonal function, given by

$$G(x, y) = \begin{cases} F(x, x, x)h(x) & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

And G is a diagonal matrix; some of the diagonal entries might be 0, so to find its rank, we need to count the number of *nonzero* diagonal entries — we have $\operatorname{rank} G = |\operatorname{supp} h| \geq |A| - m_3$.

So we have both an upper bound and a lower bound on $\operatorname{rank} G$; putting them together gives that $|A| \leq m_1 + m_2 + m_3$, which implies that the slice rank of F is $m_1 + m_2 + m_3 \geq |A|$. □

The next step is the magical step — it's where polynomials come in. (There's something intrinsically nice and mysterious about polynomials that will come up.)

Lemma 20.7

Let $F\colon A \times A \times F \to \mathbb{F}_3$ (where $A \subseteq \mathbb{F}_3^n$) be the 3-tensor given by

$$F(x,y,z) = \begin{cases} 1 & \text{if } x+y+z = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\text{slice rank}(F) \le 3 \sum_{\substack{a+b+c=n \\ b+2c \le 2n/3}} \frac{n!}{a!\,b!\,c!}.$$

Proof. Here's where we're going to use polynomials — note that for all $x \in \mathbb{F}_3$, we have

$$1 - x^2 = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \ne 0. \end{cases}$$

(You can come up with something like this for any finite field.)

So we can write $F(x,y,z)$ (where each of x, y, and z is really an element of $\mathbb{F}_3^n$) as a product of the form

$$F(x,y,z) = \prod_{i=1}^{n} (1 - (x_i + y_i + z_i)^2).$$

We devised this function looking coordinate-by-coordinate, but it's also a polynomial, which we can expand. The outcome of the expansion will be pretty messy — it'll have a lot of terms — but we can at least think about what it'll look like. It'll be some sum of terms where monomials have the form

$$x_1^{i_1} \cdots x_n^{i_n} y_1^{j_1} \cdots y_n^{j_n} z_1^{k_1} \cdots z_n^{k_n},$$

where each of the exponents is in $\{0,1,2\}$.

We want to group these monomials into three types — Type I is where $i_1 + \cdots + i_n \le \frac{2n}{3}$, Type II is where $j_1 + \cdots + j_n \le \frac{2n}{3}$, and Type III where $k_1 + \cdots + k_n \le \frac{2n}{3}$ (one of these must happen by pigeonhole).

For the contributions of the first type, we can write them as a sum of rank 1 functions by taking x as the variable we separate out — as

$$\sum_{i_1 + \cdots + i_n \le 2n/3} x_1^{i_1} \cdots x_n^{i_n} f_{i_1,\ldots,i_n}(y,z)$$

(where there may be many different exponents of y and z, but we group them all together). Each of these is a slice rank 1 function. So the number of them that we have is at most the number of summands, i.e., the number of ways to have $i_1 + \cdots + i_n \le \frac{2n}{3}$.

So to count the number of summands, we want to look at triples of nonnegative integers with $a+b+c = n$ and $b + 2c \le \frac{2n}{3}$ — here a, b, and c correspond to the number of i's which are equal to 0, 1, and 2 respectively. For each such (a,b,c), the number of corresponding summands is the number of arrangements of a, b, and c 0's, 1, and 2's. So we get

$$\text{slice rank}(F) \le 3 \sum_{\substack{a+b+c=n \\ b+2c \le 2n/3}} \frac{n!}{a!\,b!\,c!}.$$

□

Claim 20.8 — We have

$$\sum_{\substack{a+b+c=n \\ b+2c\le 2n/3}} \frac{n!}{a!\,b!\,c!} \le 2.76^n.$$

Proof. Let $x \in [0,1]$, and let's look at the expansion of

$$(1+x+x^2)^n.$$

When we expand, we'll get various contributions corresponding to each factor; and we're looking for the sum of coefficients of x^k with $k \le \frac{2n}{3}$.

But now once we have this formulation in terms of coefficients, we can see that by deleting the contributions of x^k with $k > 2n/3$ (so that $x^k \ge x^{2n/3}$, since $x \in [0,1]$), the desired sum is at most

$$\frac{(1+x+x^2)^n}{x^{2n/3}}.$$

This is some value as a function of x; and we can pick the x that minimizes it (for example, using calculus). Setting $x = 0.6$ gives that this expression is 2.76^n, which is not too far from optimal. □

(This is really the same proof as the proof of the Chernoff bounds.)

So we now have this upper bound on the slice rank of F; and we also have a lower bound from earlier. Let's put them together for the very specific function we're handed — the function that corresponds to A being a 3-AP-free set.

Proof of theorem. Let $A \subseteq \mathbb{F}_3^n$ be 3-AP-free, and define F as in the previous lemma. In a 3-AP-free set, when do we have $x+y+z=0$? This occurs exactly when (x,y,z) is a 3-AP (since $-2=+1$). If there are no 3-APs, then this shouldn't be zero unless $x=y=z$ — so

$$F(x,y,z) = \begin{cases} 1 & \text{if } x=y=z \\ 0 & \text{otherwise.} \end{cases}$$

This means it's of the form described in Lemma 4 as well as Lemma 3, so we get both a lower bound and an upper bound — combining Lemmas 3 and 4, we get

$$|A| \le \text{slice rank}(F) \le 3 \cdot (2.76)^n,$$

which gives the desired upper bound on $|A|$. □

Remark 20.9. Why is this called the *polynomial method*? The magical step is where we expand $F(x,y,z)$ as a polynomial — initially F is given discretely, and we're *writing* it as a polynomial, expanding it, and collecting terms. And this function turns out to have a very efficient polynomial form, which lets us deduce the upper bound.

This step also makes no sense in the integer setting — we know in the integer setting that you don't have an exponential improvement, but there was some hope that using ideas like this you could beat the Fourier analytic bound. But that has never been achieved.

Remark 20.10. It's an open problem to consider corners in $\mathbb{F}_3^n \times \mathbb{F}_3^n$ (i.e., the largest corner-free set) — is there an exponential improvement? The proof we saw earlier that corner-free sets have diminishing density (using the graph removal lemma) still works. But we don't know if we can get an *exponential* reduction for corners. People have tried to adapt the polynomial method proof to corners, but that has not been successful.

Remark 20.11. The same proof works in $\mathbb{F}_p^n$ for any fixed prime p. What about $\mathbb{F}_q^n$ for prime power q? The proof should also be okay, but you may have to think about what the right question is.

§20.2 Structure of set addition

We'll now move on to a different topic (though still in additive combinatorics), on the structure of set addition. We're going to be looking at *sumsets* in an abelian group. (We'll mostly think about a finite field vector space or the integers, but we can take any abelian group.)

Definition 20.12. The *sumset* of two sets A and B is $A + B = \{a + b \mid a \in A, b \in B\}$.

Note that we don't keep track of multiplicities — we just look at all possible elements of our group that can arise as a sum of an element of A and one in B.

Question 20.13. What can we say about A if $A + A$ is small?

In other words, what can we say about sets with small doubling?

First, here are some easy bounds.

Claim 20.14 — If $A \subseteq \mathbb{Z}$ has n elements, then we have

$$2n - 1 \le |A + A| \le \binom{n+1}{2}.$$

For the lower bound, we can get $|A + A| = 2n - 1$ by taking an arithmetic progression (for example $\{1, \dots, n\}$); to see that we can't do better, we can let the elements of A be $a_1 < \cdots < a_n$, and then consider $a_1 + a_1$, $a_1 + a_2$, $\dots$, $a_1 + a_n$, $a_2 + a_n$, $\dots$, $a_n + a_n$. Meanwhile, the upper bound comes from looking at all the possible combinatorial sums; the upper bound is achieved by any sufficiently generic set (for example, we can take A to consist of powers of 2).

Definition 20.15. The *doubling constant* of a set A is defined as $|A + A| / |A|$.

We'll look closer to the lower end of the above bound; in what ways can $|A + A|$ be close to its absolute minimum? The doubling constant tries to capture this relation. The word *constant* is a bit of a misnomer, but we should think of this as being bounded by some constant — that's the type of A we'll look at.

Question 20.16. What is the structure of a set A with bounded doubling constant?

For example, suppose I tell you in advance that $|A + A| \le 100\,|A|$. Can you produce some information describing the structure of A?

§20.3 Some examples

In order to even think about the answer, it's helpful to talk about some examples. What might be some ways for A to have small doubling constant?

One way is to take A to be an arithmetic progression, but there are additional examples.

Example 20.17

If A is an AP, then $|A + A| \le 2\,|A|$.

Example 20.18

If we start with an AP and delete some small fraction of its elements, then we still have roughly the same bound. We can even delete a lot more elements — if B is an AP and $A \subseteq B$ has $|A| \geq \frac{1}{k}|B|$, then

$$|A + A| \leq |B + B| \leq 2|B| \leq 2k|A|,$$

so A has doubling constant at most $2k$.

So another way to get a set with bounded doubling is by keeping a large fraction of an AP.

Another way is that instead of taking an AP (equally spaced points on a line), we can take equally spaced points on a grid — a grid also has bounded doubling. A grid is not actually a subset of $\mathbb{Z}$, but we can project onto a line using some affine transformation. Then we'll have the same additive properties in the grid and the line; so if the grid has doubling constant at most 4, then so does our line after projecting.

This is an important construction, so we'll define it and give it a name.

Definition 20.19. A *generalized arithmetic progression* (GAP) in an abelian group Γ is defined to be an affine map $\varphi\colon [L_1] \times [L_d] \to \Gamma$.

Here we view $[L_1] \times [L_d]$ as a subset of $\mathbb{Z}^d$ (a finite-sized integer grid).

Colloquially, we often refer to the *image* of φ — the set of points — as a GAP, but to be precise we also need to give the map (because given just the points, it might be ambiguous which map we mean).

By *affine* (or *affine-linear*) we mean that there exist $a_0, \ldots, a_d \in \Gamma$ such that

$$\varphi(x_1, \ldots, x_d) = a_0 + a_1 x_1 + \cdots + a_d x_d.$$

Definition 20.20. The *dimension* of a a GAP is defined as d, and its *volume* is defined as $L_1 \cdots L_d$. We say the GAP is *proper* if φ is injective.

When we take a projection, it's possible that some of our points may collide; this is generally okay, but we might not want them to, so we say a GAP is proper if they don't.

Fact 20.21 — A proper GAP of dimension d (viewed as a set) has doubling constant at most 2^d.

In d-dimensional space we're doubling in every dimension; and when we project down to the abelian group, we have the same upper bound.

Combining with our second idea gives the following example:

Example 20.22

A $\frac{1}{k}$-fraction of a d-dimensional GAP has doubling constant at most $k \cdot 2^d$.

When k and d are constants, this gives a family of constructions with bounded doubling.

§20.4 Freiman's theorem

So we've seen a few examples; we can wonder, are there more?

These types of arguments are called *direct* theorems — we're starting with some set or family of constructions, and deducing their properties.

It's much more interesting and difficult to talk about *inverse* problems — once you know the property of small doubling, what can you say about the set itself?

A cornerstone result says essentially that these are all the possible constructions.

Theorem 20.23 (Freiman 1973)

Let $A \subseteq \mathbb{Z}$ be a finite set such that $|A+A| \le k|A|$. Then A is contained in a GAP of dimension at most $d(k)$ and volume at most $f(k)|A|$, where $d(k)$ and $f(k)$ are functions only depending on k and not on A.

So in other words, ignoring the actual quantitative bounds, the construction seen above is the most general form of constructions giving sets of bounded doubling. It's an inverse theorem — knowing this property, we can recover the structure of our set.

This theorem is amazing — the property we're given seems to contain really little information, as it just says A has small doubling. But the conclusion is a very strong structural description of how A must be.

So this is a cornerstone result in the field. Its proof is quite sophisticated; it's beautiful and quite deep, and we'll spend the next several lecture going through its proof (which has many ingredients, which are interesting on their own).

§20.4.1 Abelian groups

In what sense does this generalize to abelian groups? It does generalize — this was a later result due to Green–Rusza, who generalized Freiman's theorem to general abelian groups.

First, are there other constructions that we may have missed that work in abelian groups? One possibility is that you just have a subgroup — a subgroup has no doubling. So that has to be a component.

Theorem 20.24 (Green–Rusza)

If $A \subseteq \Gamma$ is a finite set with $|A+A| \le k|A|$, then A is contained in a *coset progression* of dimension at most $d(k)$ and volume at most $f(k)|A|$.

Definition 20.25. A *coset progression* is a set of the form $P+H$ where P is a GAP and H a subgroup of Γ. The dimension of a coset progression is the dimension of P, and the volume is $\mathrm{Vol}(P) \cdot |H|$.

Remark 20.26. We mention this for cultural reasons; we will not prove it, though the proof follows similar ideas.

§20.4.2 Quantitative dependence

How do $d(k)$ and $f(k)$ depend on k? Let's think about what kinds of sets A might force our GAP to have large dimension or large volume.

What happens if we take a set with no additive relations — e.g. $2^{[n]}$ — and want to capture it by a GAP? This is pretty hard to do; the best way is to just use every element as its own dimension. So the dimension should be $n-1$. (We can of course get a set where the dimension is small, but the volume will then be really large.) In this case, we have a length-2 progression in each direction, so the volume is 2^{n-1}. And the doubling constant is around n.

We may complain because we're really supposed to think about an infinite family with bounded dimension and volume; here that's not the case. But there's a way to work around this.

How can we maintain these properties while allowing our set to grow? This is the analog of taking a blow-up of a graph — imagine our set is really long-range with no additive relations, and replace every point by an interval. Then the doubling doesn't change much — the interval just multiplies things by a factor of 2. The dimension also doesn't change much (though this requires proof), and the volume goes up proportionally.

So this gives an example of a set where the dimension and volume have to be fairly large.

We know that in some sense this is close to optimal. The proof that we'll give in the next few lectures will give much worse bounds — we'll prove bounds where $d(k) = \exp(k^{O(1)})$ and $f(k) = e^{d(k)}$. But what's known is that we can take $d(k) = k^{1+o(1)}$ (i.e., $d(k)$ can be almost linear) and $f(k) = e^{k^{1+o(1)}}$ (i.e., $f(k)$ can be almost exponential). So in some sense, this is almost optimal.

There's a separate question — if you really care about bounds, there's a better formulation of this result that says A can be *covered* by many nice objects. There's the *polynomial Freiman–Rusza conjecture*, which we'll get to later on in this chapter, which gives much better bounds by changing the problem somewhat, allowing us to *delete* some big fraction of A.

Remark 20.27. What if we replace $\mathbb{Z}$ by $\mathbb{Z}^d$? It turns out there's no difference, because of Freiman isomorphisms. But in finite field settings, there's a different story; in fact we'll begin by proving a version of Freiman's theorem in the finite field setting, where things are a lot easier.

§21 November 21, 2023

Last time we saw the statement of Freiman's theorem, a fundamental theorem in additive combinatorics giving a general description of what a set with small doubling looks like. We'll prove this theorem in the next few lectures, and on the way, we'll develop many tools central in additive combinatorics.

Our goal today is to prove a finite field version of Freiman's theorem. This will use some of the tools we'll use to prove the theorem in the integer case, but it is strictly easier (we'll need additional tools for the general result).

§21.1 Sumset calculus

The first set of tools concerns sumset inequalities. Here's an easy but important result, known as Ruzsa's triangle inequality. In this lecture, we'll generally be working with finite subsets of some abelian group (you can think of it as a finite field vector space or the integers, but the same holds for any abelian group).

Theorem 21.1 (Ruzsa's triangle inequality)
For all finite A, B, and C (which are subsets of some abelian group), we have

$$|A| \cdot |B - C| \le |A - B| \cdot |A - C|.$$

Proof. We'll construct an injection from the left-hand side (viewed as a Cartesian product) to the right-hand side. For every $d \in B - C$, we can write d as a difference of one element in B and one in C, possibly in multiple ways. We fix a way for each d — we define $\underline{b}(d)$ and $\underline{c}(d)$ to be elements from B and C, respectively, such that $\underline{b}(d) - \underline{c}(d) = d$. (In other words, we arbitrarily fix a way of writing d as a difference.)

Now we define a bijection $\varphi\colon A \times (B - C) \to (A - B) \times (A - C)$ (here $\times$ denotes the Cartesian product, not multiplication), given by

$$(a, d) \mapsto (a - \underline{b}(d), a - \underline{c}(d)).$$

To conclude the proof, we need to check that φ is injective. What it means for a map to be injective is that once we see the image, we should be able to recover where it came from. If we see the result of φ, we can obtain d by taking the difference of the two coordinates — i.e.,

$$d = (a - \underline{c}(d)) - (a - \underline{b}(d)).$$

And once we have d, we know $\underline{b}(d)$ and $\underline{c}(d)$; so then we can get a.

So we can reverse this map, which proves that it is injective. □

Remark 21.2. Why is this called a *triangle inequality*? We can define the quantity

$$\rho(A, B) = \log \frac{|A - B|}{\sqrt{|A|\,|B|}}$$

(for sets A and B in some abelian group). This quantity is known as the *Ruzsa distance*. This 'distance' satisfies the triangle inequality, in the sense that

$$\rho(B, C) \le \rho(A, B) + \rho(A, C),$$

and this inequality is precisely what we just proved (plugging in the definition). So that's the reason it's called the Ruzsa triangle inequality.

But this is purely aesthetics, in the sense that ρ is not actually a distance — to be a distance it'd also need to satisfy the property that $\rho(A, A) = 0$, and that's not true. So this is not a true distance; but it has a nice triangle inequality, which is the reason for the name.

Remark 21.3. By replacing B with $-B$ or C with $-C$, we can obtain several variants of what we just proved: replacing B by $-B$ gives that

$$|A|\,|B + C| \le |A + B|\,|A - C|\,,$$

and replacing C with $-C$ gives

$$|A|\,|B + C| \le |A - B|\,|A + C|\,.$$

Flipping both gives

$$|A|\,|B - C| \le |A + B|\,|A + C|\,.$$

(Flipping A doesn't give anything new.)

What's *missing* is a version of this inequality that says

$$|A|\,|B + C| \le |A + B|\,|A + C|\,.$$

In fact, this does not follow directly from the Ruzsa triangle inequality; but it is true, and we will prove it soon using a different set of tools.

Remark 21.4. How sharp is this inequality? You should probably be able to come up with examples where it is sharp.

§21.2 Plünnecke's inequality

Remark 21.5. A lot of theorems in this line of work — including Freiman's — have the name Ruzsa attached (for example, Freiman–Ruzsa and Plünnecke–Ruzsa). The story is that the original proofs (by Freiman and Plünnecke) were so incomprehensibly written that no one understood them, until Ruzsa came up with much simpler proofs. This is an important lesson about writing proofs clearly.

Theorem 21.6

If A is a subset of an abelian group with $|A+A| \le K|A|$, then for all nonnegative m and n, we have

$$|mA - nA| \le K^{m+n}|A|.$$

As a matter of notation, mA means A added to itself m times; this is not to be confused with scalar multiplication, which we'll denote by $m \cdot A = \{mx \mid x \in A\}$.

So Plünnecke's inequality says that if we have small doubling and we look at iterated doubling (or difference sets), we still have a bound on how big the resulting sets are — for example, if you have bounded doubling then you have bounded tripling.

Plünnecke initially proved this theorem in a bit more specialized form, with just sums and not differences. The proof used graph-theoretic ideas — specifically Menger's theorem about min-cut and flow. Ruzsa simplified the proof quite a bit — the idea is that you consider some graph (somewhat related to what we saw in the first part of the graph) with several layers, where edges correspond to adding elements of A; and you think about ways you can flow through the graph. The idea is beautiful, but the proof is quite involved; and that was the state of things for a while.

Normally, when people taught this theorem, they either bit their tongue and went through the quite involved proof, or skipped it altogether (or proved a weaker theorem that might be enough).

The situation changed quite dramatically in 2012, when Petridis (then a PhD student) came up with a radically simpler proof, which is the one we'll see — it's much shorter, but also very delicate and clever.

We will prove a slightly stronger version of the inequality.

Theorem 21.7 (Plünnecke's theorem)

If $|A+B| \le K|A|$, then for all $m, n \ge 0$ we have

$$|mB - nB| \le K^{m+n}|A|.$$

This is a bit more general — if you plug in B as A then you recover the previous result, but this statement may have its own uses.

The proof will require the following lemma, which in itself is somewhat tricky; this lemma concerns expansion ratios.

Lemma 21.8 (Expansion ratio bounds)

Let X and B be sets in some abelian group such that X is nonempty. Suppose that for all nonempty $Y \subseteq X$, we have

$$\frac{|Y+B|}{|Y|} \ge \frac{|X+B|}{|X|}.$$

Then for all nonempty (finite) sets C of the abelian group, we have

$$\frac{|X+C+B|}{|X+C|} \le \frac{|X+B|}{|X|}.$$

We'll first explain how to motivate this statement and how to think about it, in terms of expansion ratios. Consider a bipartite graph where the vertices are two copies of the abelian group, and we have edges going left to right which correspond to $+b$ for $b \in B$ — so we have a sort of Cayley graph where all the edges correspond to adding some element of the set B.

Then if we start with a set X on the left, on the right we expand X to $X + B$. And how much X expands by is the ratio we see in the statement of the lemma — so the relevant quantities are expansion ratios.

To understand the lemma, we'll first see how to use it to prove Plünnecke's theorem.

Proof of Plünnecke's theorem using Lemma. Suppose we're given that $|A + B| \le K|A|$, and we want to apply the lemma. We need to choose some X such that this hypothesis is satisfied. The hypothesis is saying that all proper subsets of X have expansion ratios at least as large as that of X; so what is a good way to choose X such that this hypothesis is automatically satisfied? We can choose X with the minimum ratio — if we make sure that the right-hand side is as small as possible, then the condition is automatically satisfied. And we're starting with A and B; so we choose a subset $X \subseteq A$ with minimum value of

$$\frac{|X + B|}{|X|}$$

(among all subsets $X \subseteq A$). Then the hypothesis of the lemma is automatically satisfied, and in particular we have

$$\frac{|X + B|}{|X|} \le \frac{|A + B|}{|A|} \le K$$

(since we chose X minimizing the left-hand side, and we could have chosen $X = A$). Now for every $n \ge 0$, if we apply the lemma with $C = nB$, we find that

$$\frac{|X + (n + 1)B|}{|x + nB|} \le \frac{|X + B|}{|X|} \le K.$$

But we can now iterate, going through all the n's — so we get that

$$|X + nB| \le K^n |X|.$$

This is quite close to what we want to prove, but so far, we've only gotten one part of it (the nB part). To get the difference, we can apply tbe Ruzsa triangle inequality — then we have

$$|mB - nB| \le \frac{|X + mB|\,|X + nB|}{|X|} \le K^{m+n}|X| \le K^{m+n}|A|$$

(since $X \subseteq A$), which is exactly what we wanted to prove. □

So we've proved Plünnecke's theorem assuming the key lemma; now it remains to prove the key lemma. The proof is pretty short, but at the time it was quite surprising the idea works — we're going to do induction on $|C|$, pulling out one element at a time. This is a pretty simple argument, but not in the style we see in additive combinatorics; the meta-lesson may be that it's good to not be deterred by the general wisdom of what works and doesn't.

Proof of Lemma. We'll do induction on $|C|$. The base case is when $|C| = 1$. In that case there is nothing to prove — $X + C + B$ and $X + C$ are just shifts of $X + B$ and X (and therefore have the same cardinalities as these sets), so the numerators and denominators on the two sides are both equal.

For the induction step, assume that we already have C with

$$\frac{|X + C + B|}{|X + C|} \le \frac{|X + B|}{|X|},$$

and let's try to add one more element γ to C — let $\gamma \notin C$, and we'll show that the same inequality holds with C replaced by $C \cup \{\gamma\}$, i.e., that

$$\frac{|X + (C \cup \{\gamma\}) + B|}{|X + (C \cup \{\gamma\})|} \le \frac{|X + B|}{|X|}.$$

The next few steps are a bit intricate. What we want to do is consider what happened when we changed the left-hand side from the old quantity to the new quantity. First, we'll look at how much the numerator increased — all the new elements are in $X + \gamma + B$. But some of these are not new elements — they could have been already present — so the change in the numerator is $|(X + \gamma + B) \setminus (X + C + B)|$. And we want to compare that to the change in the denominator, which is $|(X + \gamma) \setminus (X + C)|$ (for the same reason — we look at the new elements which came from γ and weren't already there) — it suffices to show that

$$|(X + \gamma + B) \setminus (X + C + B)| \le \frac{|X + B|}{|X|} \cdot |(X + \gamma) \setminus (X + C)|.$$

To show this, define

$$Y = \{x \in X \mid x + \gamma + B \subseteq X + C + B\}.$$

We want to look at some way to control the left-hand side in terms of Y. What's going on here? We have $|X + \gamma + B| = |X + B|$. But then we need to subtract something. What might we end up subtracting? If x is such that $x + \gamma + B$ is already entirely contained in $X + C + B$ (the thing being subtracted), then it doesn't add anything new — the translates coming from Y are going to be entirely vanished, so we can delete them from consideration, and we have

$$|(X + \gamma + B) \setminus (X + C + B)| \le |X + B| - |Y + B|.$$

(It may be helpful to think about looking at elements of X, and each gives us a translate of $\gamma + B$; and we think about which translates are present. Some are going to be entirely not present, and that's captured by the subtraction of $|Y + B|$.)

We also know that if $x \in X$ satisfies that $x + \gamma \in X + C$, then $x + \gamma + B \subseteq X + C + B$, and therefore $x \in Y$.

So one way to make sure that $x \in Y$ is to have it satisfy this (i.e., $x + \gamma \in X + C$) — it's not necessary, but it's certainly sufficient. And from this, we can deduce that

$$|(X + \gamma) \setminus (X + C)| \ge |X| - |Y|.$$

Here we're looking at which translates of γ are counted. A lot of them are going to, but some won't; which ones won't? Those are the things that satisfy $x + \gamma \in X + C$, and these are all encompassed by Y.

Now we're almost done — to prove the desired inequality (†), it suffices to show that

$$|X + B| - |Y + B| \le \frac{|X + B|}{|X|}(|X| - |Y|)$$

(simply by plugging in the two inequalities into what we were trying to show). If we look at this inequality and expand, we see the $|X + B|$ term cancels, and so this is equivalent to

$$|Y + B| \ge \frac{|X + B|}{|X|} \cdot |Y|,$$

which is true by the hypothesis on X. So that finishes the proof of the expansion ratio lemma. □

Remark 21.9. This proof is short, but it is very clever and intricate — it's quite amazing that Petridis discovered it, long after the original proof of Plünnecke's inequality, which was far more involved. The intuition here can be hard to grasp, and it's okay if we wonder how in the world anyone came up with it.

We'll now return to the missing statement in the variations of the Ruzsa triangle inequality.

Corollary 21.10
We have $|A|\,|B+C| \le |A+B|\,|A+C|$.

Proof. As before, choose (nonempty) $X \subseteq A$ with minimal $\frac{|X+B|}{|X|}$. The left-hand side involves $B+C$; we're just going to throw in an extra X, and bound $|B+C| \le |X+B+C|$. By the expansion ratio lemma, because X was chosen to satisfy the hypothesis, we see that

$$|X+B+C| \le |X+C| \cdot \frac{|X+B|}{|X|}.$$

Now $|X+C| \le |A+C|$, because $X \subseteq A$. And because X is chosen to minimize the expansion ratio, we have

$$\frac{|X+B|}{|X|} \le \frac{|A+B|}{|A|}.$$

And combining these concludes the proof. □

§21.3 Ruzsa Covering lemma

The next tool we'll introduce is a covering lemma. The idea is one that comes up in other contexts, specifically analysis.

Theorem 21.11 (Ruzsa covering lemma)
If $|X+B| \le K\,|B|$, then there exists $T \subseteq X$ with $|T| \le K$ such that $X \subseteq T+B-B$.

That's the statement; let's now explain what this is about. (The proof is pretty short, but doesn't really illustrate the basic idea, which is actually a common idea we'll see elsewhere.)

This is really a sort of geometric fact — imagine that B is a unit ball (in the plane, or Euclidean space, or some other metric space), and cardinalities are replaced by volume. Suppose we have some region X, and we want to cover X by some set of balls; and we don't want to use too many balls to cover X (but the balls can be twice as large — if B is the unit ball, we're allowed to use balls of radius 2).

So we want to efficiently cover X by balls. One way to do this is to pick a maximal set of *disjoint* unit balls with centers in X. (So instead of trying to cover X, we instead try to find a bunch of disjoint balls — we can try to put in as many as we can, or we can just try to put them in greedily until we can't put in any more.)

Now if we double these balls (expanding their radius from 1 to 2), we claim that these doubled balls cover X. This is because if they didn't cover X, then we'd be missing some point, and we could have put in another unit ball there; this would violate the maximality of our initial disjoint collection.

That's the basic idea; and you can check that you don't end up having too many balls, because of disjointness.

The idea of this lemma is essentially the same.

Proof. Let $T \subseteq X$ be a maximal subset such that the sets $t + B$ are disjoint. (The intuition is where B is a ball, but here B is any set.) Because these sets are disjoint, we see that $|T|\,|B| = |T + B| \le |X + B| \le K\,|B|$, and therefore $|T| \le K$.

To see that $X \subseteq T + B - B$, for every $x \in X$, by the maximality of T, we could not have inserted an extra copy of B at x — specifically, if we were to try to insert this extra copy $x + B$, it must intersect some existing copy $t + B$ (for some $t \in T$). So there must exist $t \in T$ with $(x + B) \cap (t + B) \neq \emptyset$. What this means is that there's a common element in $x + B$ and $t + B$, which we'll call $x + b = t + b'$. Then we have $x = t + b - b' \in T + B - B$; that finishes the claim. □

This is a pretty common technique — if you study analysis, you'll run into some very similar ideas.

§21.4 Freiman's theorem in finite fields

Now we're ready to tackle the goal stated at the beginning of the lecture — to prove a finite field analog of Freiman's theorem.

We'll first state a version over $\mathbb{F}_2$. Freiman's theorem in $\mathbb{Z}$ says that if we have a set $A \subseteq \mathbb{Z}$ with bounded doubling, then it's contained in a small GAP. The appropriate analog of a GAP is just a subspace — we want a set which intrinsically has small doubling, and subspaces just have 1-doubling (as they are closed under addition).

Theorem 21.12 (Freiman's theorem in $\mathbb{F}_2^n$)

If $A \subseteq \mathbb{F}_2^n$ has $|A + A| \le K\,|A|$, then A is contained in a subspace of cardinality at most $f(K)\,|A|$, where $f(K)$ depends only on K.

So if we have bounded doubling, then we're contained in a subspace not too much larger than the set A itself. We should appreciate that this is a nontrivial theorem — it has some content, and the conclusion shouldn't be obvious. The converse is of course much easier — if you are contained in a small subspace (i.e., if A is a large fraction of a subspace), then $A + A$ cannot exceed the subspace, so A has small doubling — so this is a converse of the statement that subspaces are closed under addition.

We'll actually prove a slightly more general form of this theorem.

Definition 21.13. The *exponent* of an abelian group (written with operation $+$) is the smallest positive integer r such that $rx = 0$ for all x in the abelian group. If no such r exists, then we say the exponent is infinite.

Remark 21.14. The final line is a matter of convention; some conventions instead say that the exponent is 0.

Example 21.15

The exponent of $\mathbb{F}_p^n$ is p, and the exponent of $\mathbb{Z}$ is infinite.

Remark 21.16. The reason this is called the *exponent* is that if we write things multiplicatively, then r instead ends up in the exponent.

We'll prove Freiman's theorem in groups with bounded exponent.

Theorem 21.17

If A is a finite subset of an abelian group with finite exponent r and $|A+A| \le K|A|$, then $|\langle A\rangle|$ (the subgroup generated by elements of A) has size at most

$$|\langle A\rangle| \le K^2 r^{K^4} |A|.$$

The point is that this constant only depends on r and K — if the exponent is bounded, this is only a function of K. This is not the optimal constant, but it's the one that we will prove.

The proof will nicely put together everything we saw today — all the different sumset inequalities and covering lemmas.

Proof. By Plünnecke's theorem, because we start with bounded doubling, we can get additional bounds on difference sets and additional iterated sumsets — in particular

$$|A + (2A - A)| = |3A - A| \le K^4 |A|.$$

(The reason for this specific expression is one we'll see momentarily; but basically because of Plünnecke's inequality we can throw in arbitrary iterated sumsets without losing anything.)

Next, by the Ruzsa covering lemma with $X = 2A - A$ and $B = A$, we find that there exists $T \subseteq 2A - A$ with $|T| \le K^4$ such that

$$2A - A \subseteq T + A - A.$$

The point of this step is that A itself could be quite large (not bounded as a function of K), but T is small (it has bounded size). And we converted $2A - A$ into something where we were able to drop the 2 and add a small set T. We can keep on iterating this to get higher iterated sumsets and bound them by some other small set plus $A - A$. Specifically, by adding A to both sides, we now get

$$3A - A \subseteq T + 2A - A.$$

(We can split the left-hand side by $A + (2A - A)$ and apply the inequality.) And then applying the inequality again, we get

$$3A - A \subseteq T + 2A - A \subseteq 2T + A - A.$$

And if we keep iterating (repeatedly adding A to both sides), we get

$$(n+1)A - A \subseteq nT + A - A.$$

Now T itself is a small set — it's fixed. In the next step we'll use the fact that we're in a group with bounded exponent, which will allow us to drop n altogether by replacing the right-hand side with $\langle T\rangle + A - A$. (In the finite field setting, $\langle T\rangle$ is the subspace generated by T; if T is a fixed set of bounded size, then this is also a set of bounded size.)

So we have

$$(n+1)A - A \subseteq \langle T\rangle + A - A$$

for all n. And in $\langle A\rangle$ (the thing we're trying to show is small), every element comes from *some* iterated sumset of A (we don't need to worry about subtraction because we can wrap around) — we have $\langle A\rangle = \bigcup_n (nA + A - A)$, and each one of these sets is contained in $\langle T\rangle + A - A$ (which doesn't depend on n), so we get

$$\langle A\rangle \subseteq \langle T\rangle + A - A.$$

And since the exponent of the group is r, we have

$$|T| \le r^{|T|} = r^{K^4}$$

(to generate this group, we put in some coefficient in front of every element of T, and the number of ways to do this is $r^{|T|}$ — some of these may be redundant, but at least this is an upper bound). Also by Plünnecke's inequality, we have

$$|A - A| \le K^2 |A|.$$

Therefore the size of the right-hand side is at most $|\langle T \rangle| |A - A| \le r^{K^4} K^2 |A - A|$, which finishes the proof. □

So that's the conclusion of this proof.

§21.5 Some comments

The goal of this chapter is to prove Freiman's theorem in $\mathbb{Z}$ (if we have a set of *integers* with small doubling, then it's contained in a generalized arithmetic progression whose size is at most a bounded multiple of $|A|$, and whose dimension is bounded).

The way we've written this proof, it doesn't extend to $\mathbb{Z}$ — there are some pretty serious difficulties. All the sumset inequalities we proved work for general abelian groups. One step that doesn't make sense for integers is when we bound $|\langle T \rangle| \le r^{|T|}$ — this really required a bounded exponent. If $T \subseteq \mathbb{Z}$, then $\langle T \rangle$ is infinite; that's not good enough.

So we'll need to come up with other ideas; and in fact we'll need to develop some new tools in the next few lectures that allow us to expand to integers.

Part of the idea is that — this is all very nice, but we also see that this is a key obstacle; and part of the reason why is that the integers ar every spread out. (They're not organized in some compact way — in $\mathbb{F}_2^n$, subspaces form much neater organization.) One thing we'll do is prove a modelling lemma that allow us to capture the essential properties of a set of integers inside a much smaller group, and then work inside that much smaller group — we'll see this idea next lecture.

Finally, we'll comment a bit about quantitative bounds, specifically for the case of $\mathbb{F}_2^n$. The proof that we saw here gavve $f(K) = 2^{K^4} K^2$. This is quite far from the truth. The exact optimum is actually known, as a function of K — it's on the order of $2^{2K}/K$. To see that this bound is tight, here's an example: consider $A = \{0, e_1, \dots, e_n\} \subseteq \mathbb{F}_2^n$. Then $A + A$ basically consists of all the vectors with support size 2, which has size roughly $\frac{1}{2}n^2$; so the expansion ratio is

$$\frac{|A + A|}{|A|} \approx \frac{n}{2}.$$

However, the smallest space containing A is the entire space, which has size 2^n; and so

$$\frac{|\langle A \rangle|}{|A|} \sim \frac{2^n}{n},$$

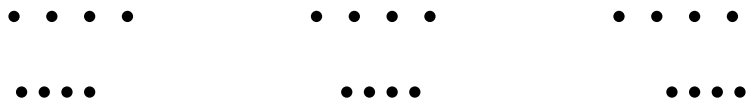
showing that this asymptotic is tight. (If you like, you can expand this example by doing a blowup — replacing every element of A by a subspace — which allows you to keep the constants but increase the size of A.)

§22 November 29, 2023

We started talking about Freiman's theorem a couple of lectures ago; the goal of this set of lectures is to prove it. Last time, we proved a version of Freiman's theorem in the finite field setting (and more generally, in a group of bounded exponent). To do so, we developed several tools (sumset inequalities and the Ruzsa covering lemma). Today we'll develop more tools that will let us prove the theorem in $\mathbb{Z}$ (which we'll do next lecture).

§22.1 Freiman homomorphisms

We've seen homomorphisms before — maps between groups that preserve the operation. For us, we're looking at sets and trying to study some of their additive structure. For example, imagine we have the following two sets A and B:



There's no way to get from A to B using a linear homomorphism, but they look very similar — structurally they're very similar, so we want to treat them as the same. For example, A and B have the same number of 3-APs. So the obvious map $A \to B$ preserves 'partial additive structure' — for example, any solution to $x + y = z + w$ is preserved (any solution in one set remains a solution in the other set).

This is a notion that we'll consider. The reason this is important is that if we only care about arithmetic progressions, then since arithmetic progressions are defined by an equation similar to this, as long as those structures are preserved we should treat the two sets as basically isomorphic.

Now let's define this notion more precisely.

Definition 22.1. Let A and B be subsets of two (possibly different) abelian groups. We say a map $\varphi\colon A \to B$ is *Freiman s-homomorphic* (or a *Freiman homomorphism of order s*) if

$$\varphi(a_1) + \cdots + \varphi(a_s) = \varphi(a'_1) + \cdots + \varphi(a'_s)$$

whenever $a_1, \ldots, a_s, a'_1, \ldots, a'_s \in A$ satisfy

$$a_1 + \cdots + a_s = a'_1 + \cdots + a'_s.$$

In words, we have a map $\varphi\colon A \to B$, and we want to say it's Freiman s-homomorphic if it preserves relations like this — if we have $2s$ numbers in a set A with the same s-wise sums, then this additive relation should be preserved under φ. Here A is not a group — it could be some small subset of a group. If φ were a genuine homomorphism, this would be automatically satisfied — because then φ distributes across addition. This condition is weaker — we just want φ to preserve these relations as long as we only care about additive relations with up to s terms.

So this is the notion of a Freiman *homomorphism*. We'd also like to define an analogous notion of a Freiman *isomorphism*.

Definition 22.2. We say φ is a *Freiman s-isomorphism* if φ is a bijection, and both φ and φ^{-1} are Freiman s-homomorphisms. Furthermore, we say A and B are *Freiman s-isomorphic* if there exists a Freiman s-isomorphism between them.

In other words, if we have two sets and all that we care about are small equations — equations like this with at most s terms on each side — then we shouldn't really need to distinguish A and B; they're isomorphic structures. This is a much weaker notion of isomorphisms than those between groups; but it's more flexible, and it'll be the one we'll use to prove Freiman's theorem.

Informally, Freiman homomorphisms (or isomorphisms) preserve s-wise sums.

Now let's discuss some properties.

- First, we have a composition structure — if φ_1 and φ_2 are both Freiman s-homomorphisms, then so is $\varphi_1 \circ \varphi_2$.

- Freiman $(s+1)$-homomorphism is automatically a Freiman s-homomorphism (and likewise with isomorphisms).
- Every abelian group homomorphism is a Freiman s-homomorphism for every s.
- If S has no nontrivial solutions to the equation $a+b=c+d$, then every map from S is a Freiman 2-homomorphism (the condition is vacuous).

Example 22.3

The obvious map $\{0,1\}^n \to (\mathbb{Z}/2\mathbb{Z})^n$ (where $\{0,1\}^n \subseteq \mathbb{Z}^n$) is a Freiman s-homomorphism for all s — because it is a group homomorphism (it's the mod 2 map on every coordinate).

But the inverse map $(\mathbb{Z}/2\mathbb{Z})^n \to \{0,1\}^n$ is not a Freiman 2-homomorphism — in $(\mathbb{Z}/2\mathbb{Z})^n$ we have

$$(1,0,0,\ldots)+(1,0,0,\ldots)=(0,0,\ldots)+(0,0,\ldots),$$

but this is not true in $\mathbb{Z}^n$.

So in particular, while the map $\{0,1\}^n \to (\mathbb{Z}/2\mathbb{Z})^n$ is a Freiman 2-homomorphism, it is not a Freiman 2-isomorphism.

A similar example holds when comparing $[N]$ (as a subset of integers) to $\mathbb{Z}/N\mathbb{Z}$. The map $[N] \to \mathbb{Z}/N\mathbb{Z}$ is a group homomorphism, so it is a Freiman homomorphism of every order; but it is not a Freiman 2-isomorphism, since its inverse is not a Freiman 2-homomorphism.

Definition 22.4. Given a set $A \subseteq \mathbb{Z}$, its *diameter* is defined as

$$\operatorname{diam}(A) = \max_{a,b\in A} |a-b|.$$

Proposition 22.5

If $A \subseteq \mathbb{Z}$ has $\operatorname{diam} A \le N/s$, then A is Freiman s-isomorphic to its image mod N in $\mathbb{Z}/N\mathbb{Z}$.

So taking the *whole* interval $[N]$ doesn't produce a Freiman isomorphism; but if we restrict A to a small *piece* of this interval, then mod N *is* a Freiman s-isomorphism.

Proof. The mod N map is a group homomorphism, so it is certainly a Freiman homomorphism. To check that it is a Freiman *s-isomorphism*, we want to show that the *inverse* map is an isomorphism — so suppose that $a_1,\ldots,a_s,a_1',\ldots,a_s' \in A$ are such that

$$(a_1+\cdots+a_s)-(a_1'+\cdots+a_s') \equiv 0 \pmod N.$$

We want to show that the same relation holds over the integers — that the left-hand side is actually 0 as an element of $\mathbb{Z}$.

To see this, we know since $\operatorname{diam}(A) < N/s$ that $|a_i - a_i'| < N/s$ for each i, so the left-hand side is strictly less than N in absolute value (as an integer). So we have something that's strictly less than N but is also 0 mod N; this means it must be 0 (as an integer). □

§22.2 Modelling lemma

In Freiman's theorem, we start with a set with small doubling. This set in some sense could be very spread out — it could be all over the place. But having small doubling intuitively says it shouldn't be so spread

out — the set itself could have very large and small numbers all over the place, but we can think of it as maybe being somewhat contained. The next lemma tries to capture that idea — if A has small doubling, we should morally think of it as being contained in a small space.

As a warmup, we'll first present a finite field version of the modelling lemma, in $\mathbb{F}_2^n$.

Proposition 22.6 (Modelling lemma in $\mathbb{F}_2^n$)

Let $A \subseteq \mathbb{F}_2^n$, and suppose that $|sA - sA| \le 2^m$ for some positive integer m. Then A is Freiman s-isomorphic to some subset of $\mathbb{F}_2^m$.

So initially n can be arbitrarily large, and A is a subset of this potentially huge ambient space. But if A starts with small doubling, then by Plünnecke's theorem we know this condition is true; and the conclusion says that even though A may potentially live in a huge space, we can model it in a much smaller space (a space of size comparable to A).

Proof. In the finite field model, things are generally a bit easier; the condition of being Freiman s-isomorphic actually has an even cleaner interpretation. In particular, the following are equivalent for a *linear* map (i.e., homomorphism) $\varphi\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$:

(1) φ is a Freiman s-isomorphism when restricted to A.

(2) φ is injective on sA.

(3) $\varphi(x) \neq 0$ for all nonzero $x \in (A - sA$.

So here we'll just be looking at linear maps, and that'll be enough — (1) is what we want, and if we can obtain this then we're done, because the image of A always sits in $\mathbb{F}_2^m$.

To see why this is the case, imagine we come up with a map φ; how could it fail to be a Freiman isomorphism? It's always a Freiman homomorphism; so it fails when we have some unintended collisions, i.e., some images end up satisfying $\varphi(a_1) + \cdots + \varphi(a_s) = \varphi(a_1') + \cdots + \varphi(a_s')$ even when we don't want them to. This gives (2) — if we look at all the s-fold sums, they should all have different images (so we shouldn't have accidental collisions).

And for (3), since φ is a linear map, so being injective on sA is equivalent to never hitting 0 on any nonzero element of $sA - sA$.

Now let φ be a linear map chosen uniformly at random. (For example, we can write down a basis and map the basis elements uniformly and independently at random; then the image of every element other than 0 will have a uniform distribution.) Then for every nonzero x, we have

$$\mathbb{P}[\varphi(x) = 0] = 2^{-m}$$

(by the uniform nature of each element). So we simply have to check whether (3) is satisfied; each x is violated with probability 2^{-m}, so by the union bound,

$$\mathbb{P}[(3) \text{ is violated}] \le \frac{|sA - sA| - 1}{2^m} < 1$$

by the hypothesis (we subtract 1 to get rid of 0). So a random map works with positive probability, proving the modelling lemma. □

Things are much simpler in the $\mathbb{F}_2^n$ case because there are lots of different linear maps, giving lots of freedom. Next we'll prove the analogous version over $\mathbb{Z}$; some of the ideas remain (we'll still be taking a random map in the proof), but you have to do more. If you're starting with a set of integers, you don't actually have so many linear maps you can work with; you have to come up with extra freedom.

So here's the statement. In fact, the statement for integers will even have a slightly weaker conclusion compared to what we just saw.

Theorem 22.7 (Ruzsa modelling lemma)

Let $A \subseteq \mathbb{Z}$, and let $s \geq 2$ and N be positive integers. Suppose that $|sA - sA| \leq N$. Then there exists $A' \subseteq A$ with $|A'| \geq \frac{1}{s}|A|$ such that A' is Freiman s-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$.

So instead of saying something about the entirety of A, we'll just say something about a large fraction of A; it turns out this will be sufficient because of other tools (in particular, the Ruzsa covering lemma).

So if we have a set with small doubling (which implies small iterated sums by Plünnecke's inequality), then we can find a pretty large portion of A — think of s as a constant (later we'll take $s = 8$ or something similar) — such that this large fraction A' is Freiman s-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$. Potentially A could have very large elements, but this modelling lemma says there's a way to compress it down to an ambient group of size comparable to the size of A (we should think of N as being on the same order as $|A|$), without losing too much additive information. So we start with something that has a lot of additive structure, and we compress it into a fairly small space (even though the set we start with might span a huge portion of $\mathbb{Z}$).

To prove this, let's take some inspiration from the finite field case. In the finite field case, we started with some set, and wanted to obtain a subset of $\mathbb{F}_2^m$; so we considered a random linear map. Here, we could *try* to consider a random map $\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$. There are a few issues with that. One is that N itself might not be a nice number; and there's not that much room for doing random maps.

But there are some ways to get around this. Instead of going directly to $\mathbb{Z}/N\mathbb{Z}$, we'll go through an intermediate setting, where there *are* a lot of random maps.

Proof. Choose a large prime q, such that $q > \max(sA - sA)$. This large prime will give us a lot of room to work with (since the cyclic group has rotations) — for every $\lambda \in [q-1]$, we can consider the map $\varphi = \varphi_\lambda$ defined as

$$\mathbb{Z} \xrightarrow{(\text{mod } q)} \mathbb{Z}/q\mathbb{Z} \xrightarrow{\cdot\lambda} \mathbb{Z}/q\mathbb{Z} \xrightarrow{(\text{mod } q)^{-1}} \{0, 1, \dots, q-1\}.$$

In words, we start with the integers, and then take mod q to get to $\mathbb{Z}/q\mathbb{Z}$. Then inside $\mathbb{Z}/q\mathbb{Z}$ we can spin things around — we can multiply by λ, which is akin to taking a random linear map in the finite field case. This takes us back to $\mathbb{Z}/q\mathbb{Z}$, and now we return to the integers — specifically $\{0, \dots, q-1\}$ — by taking the inverse of the mod q map.

The first two maps are homomorphisms; the last is not. But we saw that if we start with a set that is not too wide, then we *do* get a Freiman homomorphism — so provided that we only consider things in some interval, we'll be okay.

Since λ is chosen uniformly at random, each nonzero integer is mapped to a uniformly random element of $\{1, \dots, q-1\}$; so it is divisible by N with probability at most $1/N$. Since the number of nonzero elements in $sA - sA$ is strictly less than N, there must exist some λ such that N does *not* divide $\varphi_\lambda(x)$ for *any* nonzero $x \in sA - sA$. (This is completely analogous to the finite field setting — we check what happens for each nonzero element, and take the union bound or use linearity of expectation.)

We have an issue, that φ is not a Freiman homomorphism. But we can take care of this issue by restricting ourselves to a smaller set (this is why we have the division by s in the conclusion). By the pigeonhole principle, there exists some interval $I \subseteq \mathbb{Z}$ such that $\operatorname{diam}(I) < q/s$ and the set

$$A' = \{a \in A \mid \varphi(a) \in I\}$$

of elements in A mapped to I satisfies $|A'| \geq \frac{1}{s}|A|$.

So A starts out on the left, and it gets mapped all the way to the right (we've already fixed λ). It could end up very spread out in $\{0, \dots, q-1\}$; but we just have to pick a sub-interval where we have at least the average number of image points, and then we pull this set back to get $A' \subseteq A$. Because we've contained ourselves to a set of small diameter, what remains after restricting to A' becomes a genuine Freiman s-homomorphism.

So φ restricted to A' is a Freiman s-homomorphism (using the small diameter lemma from earlier).

We're not done yet — the goal was to get to $\mathbb{Z}/N\mathbb{Z}$. So far, we've started with a set of integers and gotten to another set of integers; we've made *some* progress, in that we managed to prevent divisibility by N. So the next thing we'll do is take mod N — let $\psi\colon \mathbb{Z} \to (\mathbb{Z}/N\mathbb{Z})$ be the map

$$\mathbb{Z} \xrightarrow{\varphi} \{0, 1, \dots, q-1\} \xrightarrow{(\text{mod } N)} \mathbb{Z}/N\mathbb{Z}.$$

This is a composition of Freiman homomorphisms, so it is a Freiman s-homomorphism; all that remains to check is that ψ sends A' Freiman-*isomorphically* to its image.

What do we need to check? Once we know something is a Freiman s-homomorphism, to check that it's a Freiman s-isomorphism intuitively means that there are no accidental collisions — φ should preserve s-wise sums, but shouldn't create more unwanted relationships. So that's what we need to check. Suppose that we have $2s$ elements $a_1, \dots, a_s, a_1', \dots, a_s' \in A'$ with

$$\psi(a_1) + \cdots + \psi(a_s) = \psi(a_1') + \cdots + \psi(a_s').$$

(This is an equality in $\mathbb{Z}/N\mathbb{Z}$.) We then want to show that the same equation without the ψ's also holds. This is the same as saying that N divides the integer

$$y = \varphi(a_1) + \cdots + \varphi(a_s) - \varphi(a_1') - \cdots - \varphi(a_s').$$

So this is what we're given, and we want to deduce that $a_1 + \cdots + a_s = a_1' + \cdots + a_s'$.

First, we can assume that y (as an integer) is nonnegative (otherwise we can swap a_i and a_i'). Since φ sent A' to something inside the interval I (which has small diameter $\operatorname{diam}(I) < q/s$), we know that $0 \le y < q$ (since y consists of s pairwise differences). So let's let

$$x = a_1 + \cdots + a_s - a_1' - \cdots - a_s' \in sA - sA.$$

(We want to show that $x = 0$.) We see that since $\varphi \bmod q$ is a group homomorphism (taking mod q in the definition of φ forgets the last map, so we just get a composition of two group homomorphisms), we must have

$$\varphi(x) \equiv \varphi(a_1) + \cdots + \varphi(a_s) - \varphi(a_1') - \cdots - \varphi(a_s') = y \pmod q.$$

We know that $\varphi(x)$ and y are congruent mod q. We want to show that they're actually the same; this follows from noting that $\varphi(x)$ and y both lie in $[0, q) \cap \mathbb{Z}$, and we just saw that they're congruent mod q; so we must have $\varphi(x) = y$.

We started with the assumption that $N \mid y$; and so now $N \mid \varphi(x)$. But we earlier chose λ so that N *never* divides $\varphi(x)$ for nonzero $x \in sA - sA$; so here we must have $x = 0$ (as an element of $\mathbb{Z}$). This means $a_1 + \cdots + a_s = a_1' + \cdots + a_s'$, which is what we wanted to prove. □

This has several steps, which may be slightly intricate (or cleverly put together). It's fine to think of this as an elaboration of the proof in the finite field case, which is conceptually much simpler — that's because the space we start with already has lots of random homomorphisms. One issue we have here is that if we start off with the given hypothesis, if N itself is e.g. a power of 2 then we don't have enough homomorphisms to spin things around. So we have to go to something nicer — a prime cyclic field, where there *are* enough nice homomorphisms to spin things around. And this creates the technical complications. But the underlying idea is the same — we work in some group where there are enough random maps to allow us to get the condition we care about using the union bound.

§22.3 Bogolyubov's lemma

Our next goal is to find a large *Bohr set* in $2A - 2A$, provided that A si a fairly large subset of $\mathbb{Z}/N\mathbb{Z}$. (We saw Bohr sets briefly when discussing Roth's theorem; we can think of them as some very structured sets analogous to subspaces in the finite field setting. In fact, we'll prove a version in the finite field setting first, where the statement is cleaner.)

Before getting to the result, we'll motivate it a bit (and explain why we need $2A - 2A$).

Question 22.8. Suppose that $A \subseteq \mathbb{F}_2^n$ is fairly large (i.e., $|A| = \alpha 2^n$ for α a constant). Must $A + A$ contain a large subspace (i.e., a subspace with codimension $O_\alpha(1)$)?

It turns out the answer is no. Here's a counterexample, called a *niveau set* (*niveau* means 'level' in French): take

$$A = \left\{ x \in \mathbb{F}_2^n \mid \operatorname{wt}(x) \le \frac{n - c\sqrt{n}}{2} \right\}$$

(where $\operatorname{wt}(x)$ denotes the Hamming weight of x, i.e., the number of 1's — so A is a level set of the Boolean lattice). By the central limit theorem, we have

$$\frac{|A|}{2^n} = \alpha + o(1)$$

where α is a positive constant depending on c. On the other hand,

$$A + A = \left\{ x \in \mathbb{F}_2^n \mid \operatorname{wt}(x) < n - c\sqrt{n} \right\}.$$

And we claim this set doesn't contain subspaces of large codimension (this is a nice linear algebraic fact that has come up before) — particularly, there is no subspace of codimension at most $c\sqrt{n}$. To see this, previously we proved that if we have a space with codimension k, then there's some element in it whose support has size at least k. So if everything here has weight at most this number, then there's no subspace with greater codimension.

So $A + A$ is not enough. But what we will prove is the following.

Lemma 22.9 (Bogolyubov's lemma in $\mathbb{F}_p^n$)
If $A \subseteq \mathbb{F}_p^n$ and $|A| = \alpha p^n > 0$, then $2A - 2A$ contains a subspace of codimension less than α^{-2}.

So if we add four copies of A instead of two, then the conclusion is true.

There's some helpful intuition here — when we add a set to itself many times, we're going to get much more regular and smooth structures (the roughness of the set evens itself out along lots of addition). This is a pretty general fact that comes up in analysis as well — suppose we start out with a real function $f\colon \mathbb{R} \to \mathbb{R}$ that's pretty jagged or rough. What happens if we convolve the function itself (taking a convolution, which is analogous to a sumset)? When we take $f * f$, this smooths out the function — that's what convolution is supposed to do — so we might get something that's a bit smoother. And if we convolve some more and look at $f * f * f$, we get something that's even smoother; and if we convolve even more, then $f * f * f * f$ is even more smooth. So more and more convolutions make your function more and more smooth. The analogy in the discrete setting is that 'smooth' means that it contains lots of nice big structures.

Let's prove Bogolyubov's analysis. We'll use Fourier analysis, so here's a reminder of some basic facts. Given $A \subseteq \mathbb{F}_p^n$ with $|A| = \alpha p^n$, recall that $\widehat{\mathbf{1}_A}(0) = \alpha$. We also have Parseval, which says that

$$\sum_r \left|\widehat{\mathbf{1}_A}(r)\right|^2 = \mathbb{E}_x \left|\mathbf{1}_A(x)\right|^2 = \alpha.$$

(In other words, the ℓ^2 norm in Fourier space is the same as the L^2 norm in physical space, which is just α.)

Proof. Let $f = \mathbf{1}_A * \mathbf{1}_A * \mathbf{1}_{-A} * \mathbf{1}_{-A}$. This function is supported on $2A-2A$ (the function $\mathbf{1}_A * \mathbf{1}_B$ is supported on the set $A+B$). Also note that

$$\widehat{\mathbf{1}_{-A}(r)} = \overline{\widehat{\mathbf{1}_A}(r)}.$$

So since the Fourier transform changes convolutions to multiplications, we have

$$\widehat{f}(r) = \widehat{\mathbf{1}_A}(r)^2 \widehat{\mathbf{1}_{-A}}(r)^2 = \left|\widehat{\mathbf{1}_A}(r)\right|^4.$$

By the Fourier inversion formula, we have that

$$f(x) = \sum_r \widehat{f}(r)\omega^{r\cdot x},$$

where $\omega = e^{2\pi i/p}$ is a pth root of unity. So plugging in our computation for $\hat{f}(r)$, we have

$$f(x) = \sum_r \left|\widehat{\mathbf{1}_A}(r)\right|^4 \omega^{r\cdot x}.$$

This allows us to write the function we care about, f, in terms of the Fourier transform of $\mathbf{1}_A$. We want to show that $2A-2A$ contains a large subspace; we'll get that subspace by thinking about how to get f to be strictly positive. We have this function f, and we want to get f to be strictly positive; and we get to pick some subspace to make that happen.

So let's pick a set of frequencies

$$R = \{r \in \mathbb{F}_p^n \setminus \{0\} \mid \left|\widehat{\mathbf{1}_A}(r) > \alpha^{3/2}\right|\}.$$

We'll basically look at the orthogonal complement of R, and that'll be our subspace. We need to show two things — first, that f is positive on $R^\perp$, and second, that R is not too large.

We can bound $|R|$ using Parseval — here we're only extracting large Fourier coefficients, and because there's a ℓ^2 bound on these coefficients, there shouldn't be too many large ones — we have

$$|R|\,\alpha^3 \le \sum_{r\in R}\left|\widehat{\mathbf{1}_A}(r)\right|^2 \le \sum_r \left|\widehat{\mathbf{1}_A}(r)\right|^2 = \alpha,$$

which means $|R| \le \alpha^{-2}$.

Now if $r \notin R \cup \{0\}$, then by definition we have $\left|\widehat{\mathbf{1}_A}(r)\right| \le \alpha^{3/2}$. The idea is that for all $x \in R^\perp$ (this is the subspace we're going to take, which has bounded codimension because $|R|$ is bounded), so that $x \cdot r = 0$ for all $r \in R$, we have by Fourier inversion that

$$f(x) = \sum_r \left|\widehat{\mathbf{1}_A}(r)\right|^4 \operatorname{Re}\omega^{r\cdot x}$$

(we can throw in a Re because $f(x)$ is real). We can split this sum into several parts, as

$$f(x) \ge \left|\widehat{\mathbf{1}_A}(0)\right|^4 + \sum_{r\in R}\left|\widehat{\mathbf{1}_A}(r)\right|^4 - \sum_{r\notin R\cup\{0\}}\left|\widehat{\mathbf{1}_A}(0)\right|^4.$$

One is the component from $r=0$, which we'll treat separately; and then elements coming from $r \in R$ all have $\omega^{r\cdot x} = 1$ (since $x \perp r$), so they contribute a strictly positive sum; and for everything else, we can naively upper bound by the absolute value.

Now we just need to show the last term is small compared to the first. The idea is to take out two of the factors and bound them by a max, and use Parseval for the rest — we have

$$\sum_{r\notin R\cup\{0\}}\left|\widehat{\mathbf{1}_A}(r)\right|^4 < \alpha^3 \sum_r \left|\widehat{\mathbf{1}_A}(r)\right|^2 = \alpha^4.$$

So then $f(x) > 0$, and therefore $R^\perp \subseteq \operatorname{supp}(f) = 2A-2A$. So this gives precisely the subspace we were looking for (which has bounded codimension because R has bounded size). □

So we found the subspace by looking at the Fourier transform, and looking at the most significant contributions to it — those gave us the desired subspace. This idea comes up repeatedly — to find structure in a set, you look at the Fourier transform or eigenvalue decomposition, and look at the directions with the largest eigenvalues, and stare at those directions and extract structure.

Now we'll do this in the integers; for that we need *Bohr sets*.

Definition 22.10. For $R \subseteq \mathbb{Z}/N\mathbb{Z}$, we define

$$\operatorname{Bohr}(R, \varepsilon) = \left\{ x \in \mathbb{Z}/N\mathbb{Z} \mid \left\| \frac{rx}{N} \right\|_{\mathbb{R}/ZZ} \le \varepsilon \text{ for all } r \in R \right\}.$$

Here rx/N sort of mimics an inner product. In the finite field setting, we want the inner product to be exactly 0; here we need a bit more leeway.

Theorem 22.11 (Bogolyubov's lemma)
If $A \subseteq \mathbb{Z}/N\mathbb{Z}$ and $|A| = \alpha N$, then $2A - 2A$ contains a Bohr set $\operatorname{Bohr}(R, \frac{1}{4})$ for some $|R| < \alpha^{-2}$.

This should be viewed as analogous to the result we just proved — finding a Bohr set with small $|R|$ is analogous to finding a subspace with small codimension.

Proof. Luckily for us, the proof is essentially verbatim — it's exactly the same as what we just did. What are the steps that need to be modified?

Here we start with $A \subseteq \mathbb{Z}/N\mathbb{Z}$, and $|A| = \alpha N$. Again we look at f being the same convolution; it's still supported on $2A - 2A$. The Fourier transform of f is also given by the exact same formula. And the Fourier inversion formula is also exactly the same, except that now $\omega = e^{2\pi i/N}$. Here R will need to change a little bit — we're looking at elements of $\mathbb{Z}/N\mathbb{Z} \setminus \{0\}$, and we still want the bound $\alpha^{3/2}$ on the Fourier coefficients. Then the inequality $|R| < \alpha^{-2}$ also remains exactly the same.

Now $R^\perp$ no longer makes sense; we need to check that for $x \in \operatorname{Bohr}(R, \frac{1}{4})$, the same result holds. The formula we use comes from the Fourier inversion formula. The contribution from $r = 0$ is still going to be the same. For r in $\operatorname{Bohr}(R, \frac{1}{4})$, we see that $rx/N \in [-\frac{1}{4}, \frac{1}{4}]$ mod 1; so $\omega^{rx} = e^{2\pi i rx/N}$ will lie on the right half of the unit circle, and so in particular its real part is nonnegative. And so we can remove the contribution from terms $r \in R$, since they're all at least 0. And we run the same bound for all the remaining terms $r \notin R \cup \{0\}$; we still get the strict bound of α^4, and so we still get $f(x) > 0$, and the conclusion still holds. □

What we've proved today is the Ruzsa modelling lemma — that a set of controlled doubling can be modelled in a fairly small set, of size comparable to the set we begin with. And then we proved Bogolyubov's lemma, which can be thought of as a continuation — once we've gotten down to a large subset of a cyclic group, we can find a large structure in $2A - 2A$.

The goal of Freiman's theorem is to show that a set of small doubling is *contained* in something with a pretty compact additive structure. This allows us to accomplish something that's spiritually similar, though it looks the other way around — it says $2A - 2A$ contains a nice structure. Next time we'll be able to convert this into GAPs, but also replace 'contains' with 'contained in.'

§23 December 4, 2023

Today we'll prove Freiman's theorem in $\mathbb{Z}$. We've been discussing it for several lectures, and we've developed many tools regarding sumsets; today we'll put them together to prove Freiman's theorem.

§23.1 Geometry of numbers

Before we do so, we'll need a few more tools — we'll need some ideas from the geometry of numbers. This is an old and classical topic that comes up in many areas (particularly number theory).

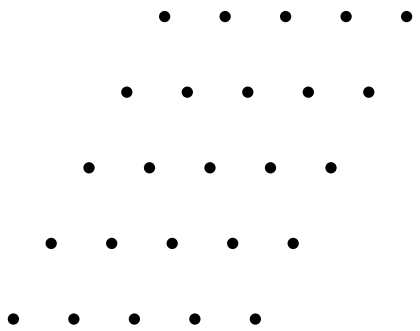
At a basic level, the study of the geometry of numbers is about lattices.

Definition 23.1. A *lattice* in $\mathbb{R}^d$ is a set of the form

$$\Lambda = \mathbb{Z}v_1 \oplus \cdots \oplus \mathbb{Z}v_d$$

where $v_1, \ldots, v_d \in \mathbb{R}^d$ are linearly independent.

Visually, a lattice in two dimensions is a set of points that looks something like this.



Given a lattice, there are various relevant quantities.

Definition 23.2. The *fundamental parallelepiped* of Λ with respect to the basis $v_1, \ldots, v_d$ is the set

$$\{x_1v_1 + \cdots + x_dv_d \mid x_1, \ldots, x_d \in [0,1]\}.$$

By translating this parallelepiped along the lattice, we get the whole space.

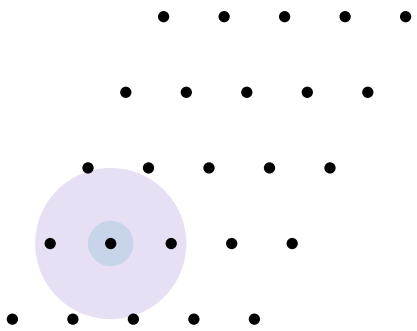
Definition 23.3. The *determinant* of Λ is defined as the determinant of the matrix of basis vectors.

Equivalently, $\det(\Lambda)$ is the volume of the fundamental parallelepiped. In a given lattice, there may be many choices for a basis to generate the fundamental parallelepiped; but all are related to each other by a matrix of determinant 1, so the determinant of this matrix does not depend on the choice of basis.

Given a lattice Λ, we can think about convex bodies on top of this lattice.

Definition 23.4. For a centrally symmetric convex body $K \subseteq \mathbb{R}^d$ (*centrally symmetric* means $x \in K$ if and only if $-x \in K$), for each $\lambda \geq 0$ we define $\lambda K = \{\lambda x \mid x \in K\}$ to be the dilation of K by a factor of λ.

You can imagine that we start with some small convex body on the origin, and then we grow it by dilation. What could happen? At some point in its growth, we might start hitting some lattice points; let's call the first time we hit this lattice point $\lambda_1 K$. And then we can record one of the basis vectors that it hits, and call that b_1.



Now we want to keep growing — as we grow, we want to find a *new* direction. If the convex body is really elongated, the next vector it sees might be $2b_1$; but we don't care, since we've already seen that. So we keep growing until we see a genuinely new direction; we call this $\lambda_2 K$, and that new direction comes with a new vector b_2. (By 'new' we mean that it isn't in the subspace spanned by the previous vectors.)

This process leads to a natural definition, namely that of a directional basis.

Definition 23.5. For Λ a lattice in $\mathbb{R}^d$ and $K \subseteq \mathbb{R}^d$ a centrally symmetric convex body, for each $1 \le i \le d$ the *ith successive minimum* of K with respect to Λ is defined to be

$$\lambda_i = \inf\{\lambda \ge 0 \mid \dim(\operatorname{Span}(\lambda K \cap \Lambda)) \ge i\}.$$

A *directional basis* of K with respect to Λ is a basis $b_1, \dots, b_d \in \Lambda$ such that $b_i \in \lambda_i K$.

So we keep growing and growing until we see i different directions; when we do, we stop and call the size λ_i. As we keep growing we'll eventually see all the directions; but during the process we record all the different stops (where we see a new direction), and we call these the λ_i. We also record these new directions, and call them a directional basis. (The directional basis consists of *lattice vectors* such that each one is contained in the corresponding dilation of K.)

So we can imagine a process where we start with K and keep growing and growing until we hit something that's a new direction. And we record this time λ_i as well as the new direction. And we keep on going; in this way we generate a set of vectors that eventually spans the entire space.

Here's an example showing some counterintuitive facts about this definition.

Example 23.6

Take the 8-dimensional lattice $\Lambda = \mathbb{Z}e_1 \oplus \cdots \oplus \mathbb{Z}e_7 \oplus \mathbb{Z}v$, where $v = \frac{1}{2}(e_1 + \cdots + e_8)$ — equivalently $\Lambda = \mathbb{Z}^8 + \{0, v\}$ consists of all points which are all integers, or all integers plus $\frac{1}{2}$.

Let K be the unit ball; what are its directional vectors?

Imagine we start with a tiny Euclidean ball at the center 0; initially it contains no other lattice points. We keep growing this ball, and eventually it's going to see some new vectors. When is the first time this happens?

The shortest vector is 1 — so we'll have $\lambda_1 = 1$. At that point, what are the lattice points contained on the sphere? We'll have $e_1, \dots, e_8$ all contained in the sphere — so the ball will contain all the coordinate vectors, and nothing else.

So what happens is that you actually get all coordinate vectors at the same time, and all the λ's equal 1 — we have $\lambda_1 = \cdots = \lambda_8 = 1$, and $b_1 = e_1, \dots, b_8 = e_8$. (We keep on growing and when we see new directions we stop and throw them in; here we stop just once and see all the directions.) And this is what this definition gives — so this is the directional basis and the successive minima.

However, the directional basis does *not* generate the point v over $\mathbb{Z}$ — so the directional basis is a basis over $\mathbb{R}$, but it is *not* a basis for the lattice Λ.

We'll need the following result, due to Minkowski.

Theorem 23.7 (Minkowski's second theorem)

Let $\Lambda \subseteq \mathbb{R}^d$ be a lattice and $K \subseteq \mathbb{R}^d$ a centrally symmetric convex body. Let $\lambda_1, \ldots, \lambda_d$ be the successive minima of K with respect to Λ. Then

$$\lambda_1 \cdots \lambda_d \operatorname{Vol}(K) \le 2^d \det \Lambda.$$

Example 23.8

When Λ is the standard integer lattice $\mathbb{Z}^d$ and $K = [-\frac{1}{\lambda_1}, \frac{1}{\lambda_1}] \times \cdots \times [-\frac{1}{\lambda_n}, \frac{1}{\lambda_n}]$ is a centrally symmetric box, this theorem is tight.

(If we imagine we have different values of λ_i we can keep growing this box; at some point we'll hit the first dimension, then the second, and so on. So the standard basis vectors get included one by one, and the λ_i actually become the successive minima.)

The proof is not long, but it's somewhat intricate (it's easy to get wrong), and we will not prove it in class (it's in the textbook, as is the statement of Minkowski's *first* theorem if we're curious).

§23.2 Finding GAPs in Bohr sets

Now let's use Minkowski's second theorem towards our goal of proving Freiman's theorem.

Last time we saw Bogolyubov's theorem, which allowed us to find in $2A - 2A$ a large Bohr set. Bohr sets are not exactly what we're looking for — we're looking for GAPs. So the next thing we'll do is go from Bohr sets to GAPs — we'll find a large GAP in a Bohr set.

Recall the following definition of a Bohr set:

Definition 23.9. Given $R \subseteq \mathbb{Z}/N\mathbb{Z}$, the *Bohr set* $\operatorname{Bohr}(R, \varepsilon)$ is defined as

$$\operatorname{Bohr}(R, \varepsilon) = \left\{ x \in \mathbb{Z}/N\mathbb{Z} \mid \left\| \frac{rx}{N} \right\|_{\mathbb{R}/\mathbb{Z}} \le \varepsilon \text{ for all } r \in R \right\}.$$

We define $|R|$ as its *dimension* and ε as its *width*.

This is meant to be an analog of subspaces (in finite field vector spaces) for $\mathbb{Z}/N\mathbb{Z}$. The next lemma allows us to get a much closer translation between these concepts.

A Bohr set is supposed to have some arithmetic structure, but the arithmetic structure we really care about is about generalized arithmetic progressions; so we'll now see how to go from one to the other.

Theorem 23.10

Let N be prime. Then every Bohr set of dimension d and width $\varepsilon \in (0, 1)$ in $\mathbb{Z}/N\mathbb{Z}$ contains a proper GAP with dimension at most d and volume at least $(\varepsilon/d)^d N$.

Definition 23.11. A *proper GAP* is the image of an injective map from some integer rectangle $[L_1] \times \cdots \times [L_d]$ to $\mathbb{Z}$. The *dimension* is d, and the volume is $L_1 \cdots L_d$.

So we start with a grid, and we get to linearly project it down to $\mathbb{Z}$ (specifying the spacing that we have); and this should be injective (that's what *proper* means). For a proper GAP, the volume is the same as size; for an improper GAP this may not be the case because of collisions.

Proof. Let $R = \{r_1, \dots, r_d\} \subseteq \mathbb{Z}/N\mathbb{Z}$. We want to be able to apply Minkowski's second theorem, so we want to set up a lattice — and the definition of the lattice is really motivated by the definition of the Bohr set. Let

$$v = \left(\frac{r_1}{N}, \frac{r_2}{N}, \dots, \frac{r_d}{N}\right) \in \mathbb{R}^d.$$

Let's see if we can rewrite the definition of the Bohr set in terms of this vector — the definition of the Bohr set is something about the entries of xv, where v is this vector and x is some scalar — it says that the coordinates of vx should all be pretty close to integers (that's why v is relevant). More precisely, for every $x = 0, 1, \dots, N-1$, we have that $x \in \mathrm{Bohr}(R, \varepsilon)$ if and only if some element of $xv + \mathbb{Z}^d$ lies in $[-\varepsilon, \varepsilon]$. In other words, if we look at xv, then mod 1, all the entries should be within ε of an integer.

This naturally leads us to consider the lattice

$$\Lambda = \mathbb{Z}^d + \mathbb{Z}v$$

generated by the standard integer lattice together with this one extra vector v. First, what is the determinant of this lattice?

Claim 23.12 — We have $\det \Lambda = \frac{1}{N}$.

Proof. The determinant of the lattice is basically the density of the lattice — it's the volume of the parallelepiped, and the more points you have per unit volume, the smaller the determinant will be.

So how many points are there in the first unit box? We basically need to count how many elements of $\mathbb{Z}v$ can be translated into the first unit box, and the answer is N — N points of Λ lie in $[0,1)^d$. (This is because v has denominator N, and N is prime; so if we think about how many translates lie in any box, the answer is N.) □

Now let's consider the convex body $K = [-\varepsilon, \varepsilon]^d$ and apply Minkowski's theorem — let $\lambda_1, \dots, \lambda_d$ be the successive minima of K with respect to Λ, and $b_1, \dots, b_d$ the directional basis.

What is the directional basis? We imagine enlarging K bit by bit until we hit the directional basis vectors and successive minima. First let's think about b_1 — what can we say about its maximum coordinate? We know that $\|b_1\|_\infty \le \lambda_1 \varepsilon$ (since we hit b_1 after dilating to $\lambda_1 \varepsilon$), and similarly $\|b_j\|_\infty \le \lambda_j \varepsilon$ for all j.

Let $L_j = \lceil 1/\lambda_j d \rceil$. Then even if we dilate b_j by some small integer, we still have $\|\ell_j b_j\|_\infty < \varepsilon/d$ as long as $\ell_j < L_j$.

So we have this lattice, which comes from starting with the integer lattice and taking this extra point v, which spins around and generates a bunch of other points. Where's the Bohr set? It corresponds to all the multiples of v that lie in a small box — v is going to spin around and get back to the origin, and some multiples of v will lie in this small box; those are going to be the vectors corresponding to our Bohr set. And we want to find a GAP associated to this structure.

What we'll do is find some really small vectors in independent directions; and they're going to span the tiny little lattice inside here. And that lattice will lie in our Bohr set, and it'll also correspond to a GAP.

Essentially, if we find ourselves a small grid in the Bohr set, then we've found a proper GAP in the Bohr set. And that's what we'll prove — the directional basis is going to be how we find this small grid, and Minkowski's second theorem will tell us that this grid actually has a lot of points.

So we've found this directional basis $b_1, \dots, b_d$. Then by the triangle inequality, we have

$$\|\ell_1 b_1 + \cdots + \ell_d b_d\|_\infty < \varepsilon$$

for all $0 \le \ell_i < L_i$. Now let's translate back to the interpretation in terms of v — for every j, there exists some $0 \le x_j < N$ such that $b_j \in x_j v + \mathbb{Z}^d$ (all the vectors in Λ are of this form, so we can represent each directional basis vector in terms of some multiple of v in this way). And then this inequality can be written in terms of these x_i's — so looking at the ith coordinate of the above inequality gives us that

$$\left\| \frac{(\ell_1 x_1 + \cdots + \ell_d x_d) r_i}{N} \right\|_{\mathbb{R}/\mathbb{Z}} \le \varepsilon$$

for all i. And that means all these values $\ell_1 x_1 + \cdots + \ell_d x_d$ are in our Bohr set — so the GAP

$$\{\ell_1 x_1 + \cdots + \ell_d x_d \mid 0 \le \ell_i < L_i\}$$

is contained in the Bohr set $\mathrm{Bohr}(R, \varepsilon)$.

We're almost done — we've found this GAP. There remains two things to check. One is that this GAP is large (which will follow from Minkowski's second theorem); and the second is that it's proper (there are no accidental collisions).

By Minkowski's second theorem, the volume of this GAP is

$$L_1 \cdots L_d \ge \frac{1}{\lambda_1 \cdots \lambda_d \cdot d^d} \ge \frac{\mathrm{Vol}(K)}{2^d \det(\lambda) d^d}.$$

We have $\det \Lambda = \frac{1}{N}$; and K is just a box. So this is

$$L_1 \cdots L_d \ge \frac{(2\varepsilon)^d}{2^d \cdot \frac{1}{N} \cdot d^d} = \left(\frac{\varepsilon}{d}\right)^d N.$$

So that gives us the volume of this GAP.

Finally, we need to check that this GAP is proper. Suppose that we have a collision — suppose that

$$\ell_1 x_1 + \cdots + \ell_d x_d \equiv \ell_1' x_1 + \cdots + \ell_d' x_d \pmod N.$$

We want to show that then the coefficients are identically equal. To do so, consider the vector b defined essentially as the vector representing the difference between theh two sides — let

$$b = (\ell_1 - \ell_1') b_1 + \cdots + (\ell_d - \ell_d') b_d.$$

On one hand, because we start with something where both sides are equal mod N, we will have $b \in \mathbb{Z}^d$. On the other hand, we have that

$$\|b\|_\infty \le \sum_{i=1}^d \frac{1}{\lambda_i \cdot d} \|b_i\|_\infty \le \varepsilon < 1$$

(since we saw that all the b_i have small entries). So b is an integer vector whose maximum entry is strictly less than 1, which means $b = 0$; and so in fact the ℓ_j and ℓ_j' are all equal, which means the GAP is proper. □

So this finishes the proof that a fairly large Bohr set contains a large GAP. We started with a Bohr set, and the idea is that we translate the condition used to define Bohr sets in terms of a lattice. Then the goal becomes roughly that we have an integer lattice and a vector v that generates an additional set of points of index N in this lattice, and we want to find many points of this lattice inside a small box which form a GAP (or rather, a small portion of a lattice). And we were able to do that by thinking about a directional basis with respect to this small box, and using a short span to generate a small lattice; and that's the small lattice that generates the GAP. And we used Minkowski's second theorem to show that this gives us enough points, enough to claim that the GAP is quite large.

Remark 23.13. Why do we say the dimension of the GAP is *at most* d rather than exactly d? It could be that some of the L_i are 1; but it'd still be okay to say that it has dimension exactly d.

§23.3 Proof of Freiman's theorem

Now we're ready to put all these tools together to prove Freiman's theorem.

First, let's restate the result.

Theorem 23.14 (Freiman's theorem)
Let $A \subseteq \mathbb{Z}$ be a finite set with $|A+A| \le K|A|$. Then A is contained in a GAP of dimension at most $d(K)$ and volume at most $f(K)|A|$, where $d(K)$ and $f(K)$ are constants only depending on A.

So a set with small doubling is contained in a small GAP.

Let's review the tools that we'll put together to prove this:

- Plünnecke's inequality, which states that small doubling implies small iterated sumsets and difference sets — specifically, if $|A+A| \le K|A|$ then $|mA-nA| \le K^{m+n}|A|$ (for all $m, n \ge 0$).
- The Ruzsa covering lemma, which states that if $|X+B| \le K|B|$, then there exists $T \subseteq X$ with $|T| \le K$ such that $X \subseteq T+B-B$. (This had a short proof where we looked at a maximal set of disjoint translates of B, and then showed that it has this property.)
- The Ruzsa modelling lemma, which says that if A is a set of integers with $|sA-sA| \le N$, then there exists $A' \subseteq A$ such that $|A'| \ge \frac{1}{s}|A|$ and A' is Freiman s-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$. So even though we may start out with a set of integers that's spread out all over the place, given that it has small doubling (and as a result, small iterated sumsets), we can compress the set (or at least, a large portion of it) to a small cyclic group without losing essential information.
- Bogolyubov's lemma, which says that for every $A \subseteq \mathbb{Z}/N\mathbb{Z}$ with $|A| = \alpha N$, the set $2A-2A$ contains some Bohr set of dimension less than $1/\alpha^2$ and width $\frac{1}{4}$ — this is a large Bohr set (the smaller the dimension, the larger the Bohr set).
- Finally, today we saw an argument (using the geometry of numbers) that a large Bohr set contains a large GAP.

Now we'll put all these nice ingredients together and prove Freiman's theorem.

Proof of Freiman's theorem. By Plünnecke's inequality, we have that $|8A-8A| \le K^{16}|A|$ (we'll see later on why we chose 8). Let N be a prime such that $K^{16}|A| \le N \le 2K^{16}|A|$ (by Bertrand's postulate there exists a prime between n and $2n$ for all n, so we can find such a prime). By the Ruzsa modelling lemma, there exists a large subset $A' \subseteq A$ — with $|A'| \ge \frac{1}{8}|A|$ — such that A' is Freiman 8-isomorphic to a subset B of $\mathbb{Z}/N\mathbb{Z}$. (So by throwing away some portion of A and keeping still a large subset, we can find a Freiman 8-isomorphic copy of this subset in $\mathbb{Z}/N\mathbb{Z}$, where N is small — on the same order as $|A|$.)

The next thing we'll do is apply Bogolyubov's lemma. To do so we need B to be a large fraction of $\mathbb{Z}/N\mathbb{Z}$, and that is true — we apply Bogolyubov's lemma with

$$\alpha = \frac{|B|}{N} = |A'|N \ge \frac{|A|}{8N} \ge \frac{1}{16K^{16}}.$$

(We can ignore all the constants and not worry too much about them; we just write them down for concreteness.) Then applying Bogolyubov, we deduce that $2B-2B$ contains a Bohr set with dimension less than $256K^{32}$ and width $\frac{1}{4}$. Then by the theorem we proved earlier today, this Bohr set in turn contains a proper GAP of dimension $d \le 256K^{32}$ and volume at least $(4d)^{-d}N$.

So we've found in $2B-2B$ a large GAP. You might wonder, is that what we really want? We want a small GAP that *contains* A; instead, we've found a large GAP that's contained *in* some model of $2A-2A$. But

it turns out these aren't so different from each other — Ruzsa's covering lemma essentially lets you go from containment to containing pretty easily, without too much loss, as we'll see now.

Since B is Freiman 8-isomorphic to A', we see that $2B - 2B$ is Freiman 2-isomorphic to $2A' - 2A'$. (If we have linear combinations of 2 terms that happen to be equal here, then they turn into linear combinations of 8 terms of B and A'.) Also note that GAPs are preserved under Freiman 2-isomorphisms (they're controlled by the equation controlling two consecutive terms — i.e., $x + y = z + z$ — which are given by Freiman 2-isomorphisms). This was the moral reason we came up with the concept of Freiman isomorphisms in the first place — to study GAPs we don't need to study all linear combinations, just ones with small coefficients.

So the proper GAP in $2B - 2B$ is mapped to a proper GAP $\mathcal{Q} \subseteq 2A' - 2A'$ with the same parameters (i.e., the same volume and dimension) — essentially, one is a fully faithful model of the other.

Now this is where we're at — instead of finding a small GAP that contains A, we've found a large GAP contained *in* $2A' - 2A'$. The final step is to convert this to what we actually want.

We have $|A| \le 8|A'| \le 8N \le 8(4d)^d |\mathcal{Q}|$ (i.e., A and $\mathcal{Q}$ have roughly the same size). Since $\mathcal{Q} \subseteq 2A - 2A$ (it's contained in $2A' - 2A'$, which is contained in $2A - 2A$), we have that $\mathcal{Q} + A \subseteq 3A - 2A$, so in particular

$$|\mathcal{Q} + A| \le K^5 |A| \le 8K^5(4d)^d |\mathcal{Q}|$$

by Plünnecke's inequality (again, all these sets have roughly the same sizes as each other — we can ignore the constants).

And now we can apply the Ruzsa covering lemma — by the Ruzsa covering lemma, there exists $X \subseteq A$ with $|X| \le 8K^5(4d)^d$ (i.e., $|X|$ is a constant) such that

$$A \subseteq X + \mathcal{Q} - \mathcal{Q}.$$

And that's kind of it — we started with something contained in A, and then applied the Ruzsa covering lemma to flip the direction of containment and cover A by a small number of translates of $\mathcal{Q} - \mathcal{Q}$.

And $\mathcal{Q}$ is a GAP, so $\mathcal{Q} - \mathcal{Q}$ is still a GAP (we can imagine taking a lattice and subtracting it from itself; that just gives a slightly larger lattice). Meanwhile X is going to give us $|X| - 1$ additional directions. So we have a GAP and we add in some constant number of additional directions; and it's still a GAP of bounded dimension (we increased the dimension by around $|X|$ and the size by around $2^{|X|}$) — so $X + \mathcal{Q} - \mathcal{Q}$ is contained in a GAP (not necessarily proper, but we don't really care about properness) with dimension at most $|X| - 1 + d = \exp(K^{O(1)})$ and volume at most $2^{|X|-1+d}K^4 |A| = \exp(\exp(K^{O(1)}))|A|$. And that's it — that proves Freiman's theorem. □

§23.4 The polynomial Freiman–Ruzsa conjecture

This proof gives us a double-exponential bound. We discussed a bit what quantitative bounds should look like. If A itself is dissociated — for example, if $A = \{10^i \mid 1 \le i \le K\}$ consists of a bunch of spread-out numbers with no additive relations — then $|A + A| \le K|A|$ (approximately). If we want to contain A in a GAP, its dimension should be roughly $K - 1$ and its volume 2^{K-1} (using a Boolean cube where each direction only has two elements). This bound is not quite what we got.

It is known that you *can* make good bounds, where $d(K) = K^{O(1)}$ and $f(K) = e^{K^{O(1)}}$ — so one can improve these bounds to something closer to the truth. The key step is a more efficient version of the Ruzsa covering lemma — instead of doing the covering all at once, if we're more careful and do it iteratively, we get a better bound. (There's a starred homework problem on this.)

In fact, we even know it's possible to get $d(K) = K^{1+o(1)}$ and $f(K) = e^{K^{1+o(1)}}$.

But the quantitative bounds bring us to another interesting question. We can even modify this example A a bit, by taking a really big base and adding a small AP — for example

$$A = \{jB^i \mid 1 \le i \le K, 1 \le j \le L\}$$

(where $B \gg L$). Then the same bounds are true (approximately).

But here this is really a combination of a small number of APs, and there's a much more efficient way to represent it — it's just a union of a small number of APs. So maybe we could get a structural description of A in this way where the bounds are much better. (It's kind of silly to try to force all the unrelated components of A into a single GAP; but if we change the theorem to allow a more flexible structure, we can hope for better bounds.)

This leads to a central problem in additive combinatorics, the polynomial Freiman–Ruzsa conjecture — a further strengthening of this result in a quantiative sense, motivated by this.

First let's discuss what happens in the finite field case.

Conjecture 23.15 (Polynomial Freiman–Ruzsa in $\mathbb{F}_2^n$) **—** If $A \subseteq \mathbb{F}_2^n$ and $|A+A| \le K|A|$, then there exists a subspace $V \subseteq \mathbb{F}_2^n$ with $|V| \le |A|$ such that A can be covered by $K^{O(1)}$ cosets of V.

Here we don't incur any exponential constants; but instead of finding a single subspace of comparable size to A containing A, we are finding a subspace that could cover A by a small number of *translates* — motivated by our example where the right way to describe the structure was not by trying to force everything into a single subspace, but rather by using a few different subspaces.

This is a version over $\mathbb{F}_2$; for a long time this was a very big open problem. Until recently, the best bound was due to Sanders, who proved a quasipolynomial bound — a bound of $e^{(\log K)^{O(1)}}$ — more than a decade ago. This was the best result for quite some time. Recently there was a huge breakthrough, where this version of the polynomial Freiman–Ruzsa conjecture was solved — this was solved about a month ago by Green–Gowers–Manners–Tao (who proved this conjecture over $\mathbb{F}_2$ — and they claim that their methods work over any prime field as well, but even $\mathbb{F}_2$ was a big open problem).

One of the nice things about this is that you can write it in this form, but there are many other equivalent forms — there are many other natural-sounding problems that turn out to be equivalent, up to a polynomial dependence on parameters, including something called the U^3 inverse theorem — so it turns out to be pretty fundamental.

There's also a version of the polynomial Freiman–Ruzsa conjecture over $\mathbb{Z}$; as of today, this is still a conjecture (i.e., it has not been proved).

We can look at the statement of Freiman's theorem and try to write down a version that only allows polynomial bounds, but allows more flexibility in how we describe A. It turns out that GAPs aren't flexible enough; we need a more flexible structure, and this is what's known as a *centered convex progression*.

Definition 23.16. A *centered convex progression* in an abelian group Γ is an affine map $\phi\colon \mathbb{Z}^d \cap B \to \Gamma$ where B is a centrally symmetric convex body; the *dimension* is d and the *volume* is $\left|B \cap \mathbb{Z}^d\right|$.

A GAP is where we start with a *box* lattice and try to map it to the integers via an affine map. But with centered convex progressions, we allow any centered convex body; this feels very similar, but we're not restricted to taking a box. Boxes are kind of arbitrary, and by allowing a more flexible but still natural set of objects, we can do a lot more.

We'll finish by stating the polynomial Freiman–Ruzsa conjecture over $\mathbb{Z}$.

Conjecture 23.17 (Polynomial Freiman–Ruzsa over $\mathbb{Z}$) — If $A \subseteq \mathbb{Z}$ has $|A+A| \le K|A|$, then one can cover A using $K^{O(1)}$ translates of some centered convex pgorgression of dimension $O(\log K)$ and volume at most $|A|$.

(Here the constant in front of $|A|$ doesn't matter — you can put a polynomial constant, with a small tradeoff.)

It takes some time to appreciate that this is the right statement — we're using only a polynomial number of translates, and this is parallel to the $\mathbb{F}_2$ version.

This is still unsolved; though given that there was a huge breakthrough in the finite field case and the community has experience translating, it's possible that this might get solved in the near future. But there may also be some difficulties with the current approach; we'll see what happens.

§24 December 6, 2023

We're going to continue our discussion of the structure of sumsets. In the past few lectures, we proved Freiman's theorem. Today we'll look at a completely different theorem, not related to Freiman's theorem but still about sumsets; it's also a very nice and important result.

§24.1 Additive energy

To describe the theorem, we need to first introduce the notion of additive energy.

Definition 24.1. For A a subset of an abelian group, its *additive energy* is

$$E(A) = \#\{(a,b,c,d) \in A^4 \mid a+b=c+d\}.$$

In other words, the energy of A counts the number of solutions to the equation $a+b=c+d$. This is some measure of additive structure — if your set has a lot of additive structure, you should expect a lot of solutions to this equation. (For example, an arithmetic progression has lots of solutions to this equation; while e.g. the powers of 2 have basically no solutions to this equation.)

In the last few lectures we considered the notion of doubling, which is a different way to measure additive structure — an arithmetic sequence has small doubling, and the powers of 2 have large doubling.

So these are two seemingly different notions of additive structure, and the goal for today is to see how they're related.

We can think of $E(A)$ as analogous to 4-cycle counts in graph theory — when we're counting solutions to this equation, we can imagine the bipartite graph where we draw edges $(x, x+a)$; then solutions to $a+b=c+d$ corresponds to going back and forth and forming a 4-cycle.

We can rewrite $E(A)$ as

$$E(A) = \sum_x r_A(x)^2,$$

where $r_A(x) = \#\{(a,b) \in A^2 \mid a+b=x\}$ is the number of pairs of elements of A that sum to x. (If we take our original expression for $E(A)$ and sum over the common sum, we get this expression.)

First, there are some easy bounds. On one hand, given three of a, b, and c, the fourth element d is automatically determined; this means $E(A) \le |A|^3$ (we have at most 3 degrees of freedom). Meanwhile, for a lower bound there are some trivial solutions — namely, where $\{a,b\} = \{c,d\}$. So this gives

$$2|A|^2 - |A| \le E(A) \le |A|^3.$$

These two bounds can basically be tight, up to constant factors — the lower bound for a completely dissociated set (e.g. powers of 2), and the other for arithmetic progressions. We would like to call sets A where the upper bound is fairly tight 'sets with additive structure' — an arithmetic progression has lots of additive structure, while powers of 2 do not.

§24.2 Small doubling vs. large energy

Question 24.2. What is the relationship between our two notions of having lots of additive structure — namely, having small doubling (i.e., $|A+A| \le K|A|$) and having large additive energy (i.e., $E(A) \ge |A|^3/K$)?

One direction is easy.

Proposition 24.3
If $|A+A| \le K|A|$, then $E(A) \ge |A|^3/K$.

So if a set has small doubling, then it has large additive energy. This intuitively makes sense — if $|A+A|$ is small, then there has to be a lot of collisions, generating lots of solutions to our equation.

Proof. We'll use the characterization $E(A) = \sum_x r_A(x)^2$, where the sum is over $x \in A+A$. By Cauchy–Schwarz, we can lower-bound this by

$$E(A) = \sum_{x \in A+A} r_A(x)^2 \ge \frac{1}{|A+A|}\left(\sum r_A(x)\right)^2.$$

Inside the parentheses we're summing over all $x \in A+A$ with multiplicity; so this gives us

$$E(A) \ge \frac{(|A|^2)^2}{|A+A|} \ge \frac{|A|^3}{K}.$$ □

On the other hand, suppose that we have $E(A) \ge |A|^3/$. Does this imply $|A+A| \le K'|A|$ (where K' is some constant depending on K)?

The answer is no. For example, take $A = [N] \cup \{2N+1, 2N+2, \dots, 2N+2^n\}$ (i.e., an arithmetic progression together with some dissociated set — we take a progression and throw in a bunch of unrelated points). This set has large additive energy — we can take the arithmetic progression and ignore the rest — but it doesn't have small doubling, since the rest of the set generates a large $|A+A|$.

So the converse doesn't hold. But here there's a big chunk of A that we'd definitely call a set with lots of additive structure. So perhaps we can get a partial converse, where in this process we may need to restrict to a large portion of A.

It turns out that's true, and that's the main theorem we'll see today.

§24.3 The BSG theorem

Theorem 24.4 (Balogh–Szemerédi–Gowers)
If $E(A) \ge |A|^3/K$, then there exists $A' \subseteq A$ with the property that

$$|A'| \ge K^{O(1)}|A| \text{ and } |A'+A'| \le K^{O(1)}|A|.$$

There's a few things worth noting here. As motivated by the above example, it's necessary to restrict to a subset $A' \subseteq A$. And we're going to lose something, but all the factors we lose are polynomial in the input parameter K. That's nice, but even qualitatively the fact that you can do this with some constant only depending on K is quite remarkable. (First Balogh–Szemerédi proved the same statement with much worse constants, potentially using the regularity lemma; then Gowers proved a much better version of the theorem.)

Gowers proved this theorem in the course of giving a new proof of Szemerédi's theorem — this was a huge revolution in the subject giving much better bounds than the previous ones. This involved many ideas, and this bound was one main step — having a bound with much better dependence on K was important.

Remark 24.5. The material we saw in the last two lectures — the proof of Freiman's theorem — and BSG both play a key role in the proof of Szemerédi's theorem for 4-APs. We saw Roth's proof for 3-AP-free sets; the 4-AP-free proof is much more involved and involves many ideas, and in particular we need to better understand additive structure; this result and Freiman's theorem are some of the key ingredients.

The goal of the rest of this lecture is to prove the BSG theorem. First we'll state a variant of this theorem, the one we'll actually prove. To state this, we'll need to define additive energy between two *different* sets.

Definition 24.6. For two sets A and B, we define $E(A,B) = \{(a,b,a',b') \in A \times B \times A' \times B' \mid a+b = a'+b'\}$.

Theorem 24.7 (BSG version 2)
If $|A|, |B| \le n$ and $E(A,B) \ge n^3/K$, then there exist $A' \subseteq A$ and $B' \subseteq B$ such that

$$|A'| \ge K^{-O(1)}|A|, |B'| \ge K^{-O(1)}|B|, \text{ and } |A'+B'| \le K^{O(1)}n.$$

First, what's the relationship between these two statements? It's not obvious whether the first implies the second, but we'll show the second implies the first. It's not completely trivial — you could try taking $A = B$, but we'll get A' and B' in the conclusion, and we just want a single A'. So there's work to do, but we can do this.

Proof BSG2 $\implies$ BSG1. Applying BSG2 with $A = B$ gives two potentially different sets $A', B' \subseteq A$. Now we can apply the Ruzsa triangle inequality to upper-bound $|A'+A'|$ — we have

$$|A'+A'| \le \frac{|A'+B'|^2}{|B'|}.$$

(This is the slightly harder version of the Ruzsa triangle inequality, but we did prove it.) The BSG2 conclusion upper-bounds $|A'+B'| \le K^{O(1)}|A'|$, so we're done. □

The message to take away is that if we have some nice hypotheses, then these small differences don't matter too much — if we can understand $A'+B'$, then we can also understand $A'+A'$.

§24.4 A graph version

We will prove BSG2 using graph theory — we will set up a graph that allows us to prove this result.

Definition 24.8. Suppose we have sets A and B in an abelian group, and G is a graph with vertex bipartition $A \cup B$. We define the *restricted sumset* $A +_G B$ to be the set

$$A +_G B = \{a + b \mid (a, b) \in A \times B \text{ is an edge in } G\}.$$

So we have some A and B, and some graph in between them, and we're only allowed to add along edges. In particular, if G is a complete graph then this is the usual notion of sumsets; but here we are only allowed to add along edges of G.

We will state a graphical version of the BSG theorem (which we'll then show implies the BSG theorem).

Theorem 24.9 (Graph BSG)
If $|A|, |B| \le n$ and $e(G) \ge n^2/K$, and $|A +_G B| \le Kn$, then there exist $A' \subseteq A$ and $B' \subseteq B$ such that

$$|A'|, |B'| \ge K^{-O(1)} n \text{ and } |A' + B'| \le K^{O(1)} n.$$

So we're given that a restricted sumset of A and B over a graph with reasonably many edges is fairly small; and we get to conclude that we can find subsets whose *unrestricted* sumset is small. Here there's no notion of additive energy — instead, we say that if we have a small sumset along a dense graph, then we have a small complete sumset when restricted to some large sets.

Proof Graph BSG $\implies$ BSG2. As with $r_A(x)$, define $r_{A,B}(x) = \#\{(a, b) \in A \times B \mid x = a + b\}$ be the number of ways to write x as $a + b$. In BSG2, we start out with the hypothesis of having large energy; somehow we need to convert it to the notion of having a small restricted sumset along some graph. So which graph should we take? We'll take a *popular sums* graph — the graph will record sums that are popular.

We define the set of *popular sums* as

$$S = \left\{ x \in A + B \mid r_{A,B}(x) \ge \frac{n}{2K} \right\}.$$

(So popular sums are those that can be written in lots of ways as $a + b$.) Once we have S, we build the bipartite graph G between A and B such that $(a, b) \in A \times B$ is an edge if and only if $a + b \in S$ — so we only keep the edges that correspond to popular sums.

Why is it a good idea to do this? By the assumption that we have large additive energy, the popular sums have to contribute a lot to the energy — you have to have lots of popular sums. And in fact (as we'll see now), by only keeping the popular sums, you still generate lots of energy; so you don't lose much by restricting to popular sums.

Claim 24.10 — Unpopular sums account for less than $\frac{1}{2}$ of $E(A, B)$.

Proof. We have $E(A, B) = \sum_x r_{A,B}(x)^2$. We can split this sum into popular sums and unpopular sums, as

$$E(A, B) = \sum_{x \in S} r_{A,B}(x)^2 + \sum_{x \notin S} r_{A,B}(x)^2.$$

For the unpopular sums, by definition we have

$$\sum_{x \in S} r_{A,B}(x)^2 \le \frac{n}{2k} \cdot |A|\,|B| \le \frac{n^3}{2K}.$$

This means the unpopular sums in total contribute at most half the energy, as desired. □

So the popular sums contribute at least half the energy, which means the number of popular sums and the number of edges in G must be fairly large — we have

$$e(G) = \sum_{x \in S} r_{A,B}(x) \geq \sum_{x \in S} \frac{r_{A,B}(x)^2}{n} \geq \frac{n^2}{2K}$$

(since of course $r_{A,B}(x) \leq n$). Furthermore, $|A +_G B|$ has to be small — each element corresponds to some popular sum, so it represents at least $n/2K$ edges. But the total number of edges is at most $|A|\,|B| \leq n^2$, so we have

$$\frac{n}{2K} |A +_G B| \leq |A|\,|B| \leq n^2,$$

and therefore $|A +_G G| \leq 2Kn$.

That proves that graph BSG implies BSG2. □

Remark 24.11. This proof may take some time to digest, but it's kind of a routine computation that we often see in extremal and probabilistic combinatorics — there may be some small insignificant things happening, but we throw them out and focus on the regular part of the structure where interesting things happen, where things are easier to analyze. We'll see this many times in our proof of graph BSG in the rest of this lecture.

Remark 24.12. Are the constants reasonable? They should be; they're probably less than 20 if you're careful. (But the optimal exponent may be something quite small, like 5 or 6.)

Remark 24.13. In additive combinatorics, there's lots of steps that look like the conclusion we want here, where we keep passing down to subsets. In fact, when we proved Freiman's theorem, there was a step in Ruzsa's modelling lemma where we had to pass down to a subset. Often that's not too bad — the covering lemma allows you to go from covering a subset to the whole thing.

In the rest of this lecture, we'll prove graph BSG, in several steps.

§24.5 Some path lemmas

First we'll need the path of length 2 lemma. Informally, it says that if we have two sets A and B and lots of edges between them, then we can find a fairly large set $U \subseteq A$ such that every pair of vertices in U have many common neighbors.

Lemma 24.14 (Path of length 2 lemma)

Let $\delta, \varepsilon > 0$ and let G be a bipartite graph on $A \cup B$ with $e(G) \geq \delta\,|A|\,|B|$. Then there exists $U \subseteq A$ with $|U| \geq \frac{\delta}{2}\,|A|$ such that at least a $(1-\varepsilon)$-fraction of pairs $(x, y) \in U^2$ have at least $\frac{1}{2}\varepsilon\delta^2\,|B|$ neighbors common to x and y.

So we have a fairly dense graph, and the conclusion is that there must exist a large subset U on the left such that for almost all pairs (x, y) in U, the number of common neighbors of x and y is pretty large (a constant fraction of B).

We saw something like this much earlier, with dependent random choice (we saw this in a more general form — if we want to get a large subset U such that every r elements in U have many common neighbors, then we can do it). This is an even easier version, where we only require 2 vertices and 'almost all,' but the idea is the same; we'll see the proof again.

Proof by dependent random choice. We don't want to select U randomly in A; that's not going to work. So instead, we select something random in B, and create U out of the randomness in B. In this specific case, we select U to be the neighborhood of a random vertex in B.

To do the analysis, we say a pair $(x, y) \in A^2$ is *unfriendly* if x and y have very few common neighbors — specifically, fewer than $\frac{1}{2}\varepsilon\delta^2 |B|$ common neighbors. We want to restrict to some set such that there are very few unfriendly pairs.

First, here's the randomness: we choose a vertex $v \in B$ uniformly at random, and let $U = N(v)$. So U comes from choosing v on the right, and looking at its neighborhood on the left.

Then $\mathbb{E}|U| = \mathbb{E}|N(v)|$, which we can find by linearity of expectation — the average neighborhood size is the same as

$$\mathbb{E}|N(V)| = \frac{e(G)}{|B|} \geq \delta |A|.$$

So U in expectation has large size.

Meanwhile, for all fixed $(x, y) \in A^2$ (note that x and y are not random), we have

$$\mathbb{P}(x, y \in U) = \mathbb{P}(x \sim v, y \sim v) = \frac{\operatorname{codeg}(x, y)}{|B|}$$

(since x and y both end up in U if and only if they're both neighbors of v). If x and y are unfriendly, then this quantity is strictly less than $\varepsilon\delta^2/2$. This tells us unfriendly pairs are unlikely to both be in U at the same time.

This is a good sign — typically it seems that in expectation U is pretty large, and U should have a small number of unfriendly pairs (both of which are what we want). Then we'll do a routine argument to convert to the *existence* of such a U for which both are true.

Let X be the number of unfriendly pairs in U. Then we have $\mathbb{E}X < \frac{\varepsilon\delta^2}{2}|A|^2$, so by convexity

$$\mathbb{E}\left[|U|^2 - \frac{|X|}{\varepsilon}\right] \geq (\mathbb{E}|U|)^2 - \frac{\mathbb{E}X}{\varepsilon} > \frac{\delta^2}{2}|A|.$$

So there exists v such that setting $U = N(v)$, we have

$$|U|^2 - \frac{X}{\varepsilon} \geq \frac{\delta^2}{2}|A|^2.$$

And if we have this inequality, then this U definitely works.

This is basically the same as the proof of dependent random choice, except without the extra step of making sure *every* pair has lots of common neighbors. □

That was the path of length 2 lemma; now let's upgrade it to a path of length 3 lemma.

The idea is that we basically want to say that given A and B, we want to restrict to large A' and B' such that for each $a \in A'$ and $b \in B'$, there are lots of paths of length 3 between them.

Lemma 24.15 (Path of length 3 lemma)

Let $\delta > 0$, and let G be a bipartite graph on $A \cup B$ with $e(G) \geq \delta |A||B|$. Then there exist $A' \subseteq A$ and $B' \subseteq B$ such that

$$|A'| \geq c\delta^C |A| \text{ and } |B'| \geq c\delta^C |B|$$

and such that the number of 3-edge paths joining every pair $(a, b) \in A' \times B'$ is at least $c\delta^C |A||B|$.

(We use the convention that c represents a small constant and C a large constant.)

We'll prove this using the path of length 2 lemma. We have two sets A and B. There'll be many steps in this proof where we want to get rid of low-degree vertices (because they're annoying and contribute very little); so there'll be many steps like this. In the course of this sketch, we can pretend to not see these insignificant contributions.

Proof. We start with A and B. First we can clean A up a bit, and then apply the path of length 2 lemma to get a large subset of A with the property that every pair of vertices has lots of common neighbors.

That's not quite enough — the structure we're looking for is a path of length 3. But what if we have for free that every vertex in A has lots of outgoing edges (i.e., A has large minimum degree)? We can also for now imagine that things are distributed fairly regularly. Suppose we can find A', and then find B' such that every vertex in B' has lots of common neighbors to A'. Then for any $a \in A'$ and $b \in B'$, we can first start at b, and because we have lots of neighbors in A' we can take a first step in many ways. And then we can do a path of length 2 to complete this 3-path.

We know how to find A', because of the path of length 2 lemma. But how do we find B'? We want the property that every vertex in B' has lots of neighbors in A'.

First, something slightly easier: is it easy to get A' such that every vertex in A' has lots of neighbors going right? One way to get this is to just throw away, at the beginning, all the disqualifying vertices — throw away all the low-degree vertices in the beginning, and then we don't have to worry anymore about low-degree vertices in A. Then we have the situation that every vertex in A' has lots of outgoing edges to the right.

And then the graph between A' and B has lots of edges between them, so we can do the same thing, throwing away vertices in B with low degree to A'. And then whatever remains is going to have large degree to A'.

Remark 24.16. If you execute this properly, you don't have to go back and forth.

So this is the basic idea that comes up a lot in combinatorics — you have something that starts off fairly irregular, and you do a number of cleaning steps to make it regular. These cleaning steps are often quite messy or annoying and not the most fun to read, but they're also not hard. We'll be light on the details, but the idea is basically what's described — we throw away insignificant contributions to get something regular, in which case we can deduce the result.

So the idea is to repeatedly trim low-degree vertices. We fisrt restrict A to

$$A_1 = \left\{ a \in A \mid \deg a \ge \frac{\delta}{2} |B| \right\}.$$

(The average degree is $\delta |B|$.) Then we can check that the number of edges that remain after this restriction is still a positive fraction of the original number of edges — specifically

$$e(A_1, B) \ge \frac{\delta}{2} |A| |B|$$

(we started with $\delta |A| |B|$ and threw away at most $\frac{\delta}{2} |A| |B|$); this automatically implies $|A_1| \ge \frac{\delta}{2} |A|$.

Next we construct $A_2 \subseteq A_1$ by the path of length 2 lemma, such that $|A_2| \ge \frac{\delta}{2} |A_1| \ge \frac{\delta^2}{4} |A|$ and all but a $\frac{\delta}{10}$-fraction of pairs of vertices in A_2 have at least $\frac{\delta^3}{20} |B|$ common neighbors. (We should ignore the precise expressions of the constants; it only matters that they are polynomial in δ.) So we first deleted some insignificant vertices, and then passed down using the path of length 2 lemma.

Now we go to the B-side and do some restrictions there — we set

$$B' = \left\{ b \in B \mid \deg(b, A_2) \ge \frac{\delta}{4} |A_2| \right\}.$$

(We know after applying the path of length 2 lemma we have lots of edges between A_2 and B, and we restrict down to get rid of low-degree vertices on the right.) Then we can show that $e(A_2, B') \geq \frac{\delta}{4} |A_2| |B|$ (skipping a few details — we can check that the contributions to $e(A_2, B)$ from the deleted vertices was pretty small, since we only deleted low-degree vertices), so in particular $|B'| \geq \frac{\delta}{4} |B|$ is pretty large.

Finally, the path of length 2 lemma tells us that *almost* all pairs have lots of common neighbors. If we actually had 'all' instead of 'almost all' then we'd already be done. But right now if we start with some $a \in A$ we might get unlucky and not be able to go to many other vertices. This isn't really a problem; we can again get around this by neglecting a small number of vertices. So let

$$A' = \{a \in A_2 \mid a \text{ is friendly to at least a } (1 - \frac{\delta}{3})\text{-fraction of } A_2\}$$

(where 'friendly' means having at least $\frac{\delta^3}{20} |B|$ common neighbors). We can check that $|A'| \geq \frac{|A_2|}{2} \geq \frac{\delta^2}{8} |A|$.

We claim that A' and B' do the job. The reason is basically what we said earlier — if we have some vertex $a \in A'$ and $b \in B'$, we want to know there are many paths of length 3 between them. The vertex b has lots of neighbors to A — it has at least a $\frac{\delta}{4}$-fraction of neighbors in A_2. And a is unfriendly to a small number of other vertices, but less than $\frac{1}{5}\delta$ of A_2; so there'll be some portion here that a is friendly to, and for each of them we get lots of paths of length 2. Then we can count all the way through, and we get the result we want. □

§24.6 Proof of graph BSG

The final step is to actually prove the graph BSG theorem, and that's what we'll do now.

Proof. Since $e(G) \geq n^2/K$, we have that $|A|, |B| \geq n/K$. (They're at most n, so if they're too small we can't get this many edges to begin with.)

So by the path of length 3 lemma, we can find $A' \subseteq A$ and $B' \subseteq B$, each with large size — meaning $|A'|, |G'| \geq K^{-O(1)}n$ — such that for all $(a, b) \in A' \times B'$ there exist lots of paths (a, b_1, a_1, b) in G (where by 'lots' we mean at least $K^{-O(1)}n^2$).

Let's see how this helps us. We'll define a few auxiliary variables — let $x = a + b_1$, $y = a_1 + b_1$, and $z = a_1 + b$. (These are the sums along the edges in our path of length 3.) From these algebraic relations, we have

$$a + b = x - y + z$$

(this is a key identity in this proof). So every element of $A' + B'$ can be written as $x - y + z$ for some x, y, and z in the *restricted* sumset $A +_G B$ in lots of ways — i.e., at least $K^{-O(1)}n^2$ ways. (There are lots of length-3 paths, and each gives such a decomposition of $a + b$.) Furthermore, for each fixed $(a, b) \in A' \times B'$, these choices of x, y, and z are distinct (i.e., these paths correspond to distinct elements of $A +_G B$ — because they explain how to travel from one vertex to the other).

So now $|A' + B'|$ is the number of sums $a + b$, and each of these sums corresponds to at least $K^{-O(1)}n^2$ triples (x, y, z); this gives

$$K^{-O(1)}n^2 |A' + B'| \leq |A +_G B|^3 .$$

(For each sum $a + b$ we can represent it with (x, y, z) in at least $K^{-O(1)}n^2$ ways, but the right-hand side bounds the total number of choices of x, y, and z.)

But the hypothesis of graph BSG is that $|A +_G B| \leq Kn$, and rearranging gives

$$|A' + B'| \leq \frac{K^3 n^3}{K^{-O(1)}n^2} \leq K^{O(1)}n,$$

finishing the proof of graph BSG. □

This completes our discussion of sumsets; we won't have time to discuss it in this class, but this result combined with Freiman's theorem are central results in additive combinatorics, and part of the reason is that they feed into the proof of Szemerédi's theorem for longer progressions (in particular 4-term progressions), where one really has to understand the structure of sumsets.

§25 December 11, 2023

Today we'll talk about the sum-product problem.

§25.1 The sum-product problem

So far, we've discussed some core problems in additive combinatorics — Roth's theorem and the structure of sumsets. These only involved additive structure. Today we'll look at additive structure together with multiplicative structure.

Definition 25.1. We define $A+A = \{a+b \mid a,b \in A\}$. Similarly we define $AA = A\cdot A = \{ab \mid a,b \in A\}$.

If we only look at AA, this is the same as looking at the sumset of the logarithms. But when we look at both at the same time, new interesting problems emerge.

Question 25.2. Can both $A+A$ and $A\cdot A$ be small at the same time?

We know $A+A$ is small for an arithetic progression, whereas $A\cdot A$ is small for a *geometric* progression — and arithmetic progressions and geometric progressions look very different. For example, the powers of 2 have very large sumset. These don't seem to be compatible with each other; and we believe that it's *not* possible to make both small at the same time.

Conjecture 25.3 (Sum-product conjecture) — We have $\max\{|A+A|, |A\cdot A|\} geq |A|^{2-o(1)}$.

This is due to Erdős and Szemerédi; it says that at least one of the sumset and product set should be essentially as large as a dissociated set. (This is very much open.)

Erdős and Szemerédi (in 1983) proved that $\max\{|A+A|, |A\cdot A|\} \geq |A|^{1+c}$. Today we'll see two proofs that give different values of c, but the optimal c is conjectured to be 1, and we don't have a proof of that.

Question 25.4 (Erdős multiplication problem)**.** What is $|[N]\cdot[N]|$?

In other words, how many distinct numbers appear in the $N\times N$ multiplication table? It's actually a bit less than N^2. This is a problem in analytic number theory, and we actually now know the answer quite precisely — a result due to Ford (2008) shows that

$$|[N]\cdot[N]| = \Theta\left(\frac{N^2}{(\log N)^\delta (\log\log N)^{3/2}}\right)$$

where δ is a specific constant (whose value we know). We won't discuss where this comes from (Prof. Zhao doesn't know either); but in the textbook, there is a short proof that $|[N]\cdot[N]| = o(N^2)$ (this is already interesting and nontrivial) and $|[N]\cdot[N]| \geq (1-o(1))N^2/2\log N$ (so it's not too much smaller than N^2). This in particular shows why we need the $o(1)$ in the sum-product conjecture.

§25.2 The first proof

The first proof we'll give, due to Elekes, proves the following bound.

Theorem 25.5 (Elekes)
Every $A \subseteq \mathbb{R}$ satisfies $|A+A|\,|A\cdot A| \gtrsim |A|^{5/2}$.

This gives that $\max\{|A+A|, |A\cdot A|\} \geq |A|^{5/4}$.

This will be a very nice proof allowing us to take a detour back to graph theory. We'll use a result in graph theory known as the crossing number inequality.

§25.2.1 The crossing number inequality

Definition 25.6. The crossing number of a graph G, written $\mathrm{cr}(G)$, is the minimum number of edge crossings in a planar drawing of G where edges are drawn as continuous curves.

Given a grpah, you can try to draw it in the plane. If the graph is planar you don't need any crossings, but if it's not planar you will need some. There's a subtlety in this definition — you can have variants such as requiring straight lines as edges, or you can count pairs of edges that cross. These definitions are not exactly the same, but we won't get into that; those fine details aren't so important for what we'll discuss.

Question 25.7. Given a graph with lots of edges, does it necessarily have lots of crossings?

In particular, suppose the graph has quadratically many edges; how many crossings does it have to have?

The crossing number inequality addresses that; it's due to two independent papers from the 1980s.

Theorem 25.8 (Crossing number inequality)
Every graph $G = (V, E)$ with $|E| \geq 4\,|V|$ has

$$\mathrm{cr}(G) \gtrsim \frac{|E|^3}{|V|^2}.$$

Imagine that $|E|$ is on the order of n^2 (where $|V| = n$). Then this bound is on the order of n^4. This is also a trivial upper bound — the number of pairs of edges is $\Theta(n^4)$. So this theorem says that lots of edges implies lots of crossings.

Proof. We'll first prove this by some basic facts about planar graphs.

Fact 25.9 — Every connected planar graph $G = (V, E)$ with at least one cycle has $3\,|F| \leq 2\,|E|$.

Proof. We can count edges by going through all the faces; each face has at least 3 edges, and each edge is counted twice. □

Then Euler's formula states that

$$|V| - |E| + |F| = 2,$$

and replacing $|F|$ with what's given by the above inequality and rearranging gives that

$$|E| \leq 3\,|V| - 6.$$

This is true for every connected planar graph with at least one cycle; and from this we can deduce that for *every* planar graph G (dropping the connected and at least one cycle hypotheses) we have $|E| \le 3|V|$. (This is true without cycles, and we can then apply this to every component.)

Now if G is a graph with $|E| > 3|V|$, then any drawing of G can be made planar (i.e., can be made to have no crossing edges) by deleting at most $\mathrm{cr}(G)$ edges. (There's some crossings; we delete them and then we have a planar drawing.) Then the graph that has been made planar satisfies

$$|E| - \mathrm{cr}(G) \le 3|V|$$

(by our above inequality — the new graph has at least $|E| - \mathrm{cr}(G)$ edges.) So this means

$$\mathrm{cr}(G) \ge |E| - 3|V|.$$

This is a lower bound on $\mathrm{cr}(G)$, which is in the direction we're looking for; but it's not the order of magnitude we want (if we have n^2 edges, this only gives us a bound on the order of n^2, but we want one on the order of n^4).

So the next idea is to boost this bound to something better, which we'll do using a probabilistic method. The idea is that we pick some $p \in [0,1]$, whose precise value we will decide later, and let $G' = (V', E')$ be the graph obtained by keeping every vertex of G with probability p. (So we delete a bunch of vertices at random.) Now G' is still a graph, and so the weak lower bound still holds for G' — we have

$$\mathrm{cr}(G') \ge |E'| - 3|V'|.$$

(This is true for every G', so it's in particular true for our random one.) But then we have

$$\mathbb{E}[\mathrm{cr}(G')] \ge \mathbb{E}[|E'|] - 3\mathbb{E}[|V'|].$$

But $\mathbb{E}[|V'|] = p|V|$, and $\mathbb{E}[|E|] = p^2|E|$ (since an edge is kept if both its endpoints are kept).

For $\mathbb{E}[\mathrm{cr}(G')]$, we could have kept the original drawing of G, in which case each crossing is kept with probability p^4 — so we have $\mathbb{E}[\mathrm{cr}(G')] \le p^4 \mathrm{cr}(G)$ (it could be smaller, since there could be a better drawing, but this inequality is true). So we have

$$p^4 \mathrm{cr}(G) \ge p^2|E| - 3p|V|.$$

Now we can optimize a value of p — let $p = 4|V|/|E|$ (the reason we need the hypothesis is to ensure that $p \in [0,1]$). Then after doing some algebra, we get

$$\mathrm{cr}(G) \ge p^{-2}|E| - 3p^{-3}|V| \gtrsim \frac{|E|^3}{|V|^2}$$

(the point of choosing a godo value of p is essentially to ensure that the second term is a bit smaller than the first). □

To recap, we first take a basic inequality for planar graphs, use it to prove a weak lower bound for $\mathrm{cr}(G)$ for all graphs, and use the probabilistic method to boost to a much better bound. (The constant of 4 can be replaced by anything greater than 3, but can't be made smaller than that, since there exist planar graphs with $|E| \approx 3|V|$ — for example, take a large hexagonal grid.)

Remark 25.10. For planar graphs, it doesn't matter whether you require straight lines or not. (A graph is planar if and only if you can draw it using straight lines — in fact, a graph is planar if and only if you can draw it using representations of circles.) In general, there are many subtle variations of crossing numbers (where we allow straight lines, or count with multiplicity when two curves cross many times), and it is a very subtle question of whether they are the same.

§25.2.2 Point-line incidences

We'll now use the crossing number inequality to prove an important result about point-line incidences.

Definition 25.11. Given a (finite) set of points $\mathcal{P}$ and a set of lines $\mathcal{L}$ in the plane, we define

$$\mathcal{I}(\mathcal{P},\mathcal{L}) = \{(p,\ell) \in \mathcal{P}\times\mathcal{L} \mid p \in \ell\}.$$

(In other words, we count the number of pairs of a point lying on a line.)

Question 25.12. What is the maximum possible number of incidences between n points and m lines?

This turns out to be quite a subtle question, although now we understand it up to a constant factor.

First, there's a trivial bound of $|\mathcal{I}(\mathcal{P},\mathcal{L})| \le |\mathcal{P}|\,|\mathcal{L}|$. We can do better by noting that every pair of points determines at most one line. So using this alone, we can notice that

$$|\mathcal{P}|^2 \ge \#\{(p,p',\ell) \in \mathcal{P}\times\mathcal{P}\times\mathcal{L} \mid pp' = \ell \text{ and } p \ne p'\}$$

(since two points determine a line). On the other hand, we can count this set by summing over all lines ℓ, and choosing two distinct points on each line — so this is equal to

$$\sum_{\ell\in\mathcal{L}} |\mathcal{P}\cap\ell|\,(|\mathcal{P}\cap\ell|-1).$$

We can lower-bound this by Cauchy–Schwarz to get that

$$|\mathcal{P}|^2 \ge \frac{|\mathcal{I}(\mathcal{P},\mathcal{L})|^2}{|\mathcal{L}|} - |\mathcal{I}(\mathcal{P},\mathcal{L})|\,.$$

That's an inequality, and by rearranging it we find that

$$|\mathcal{I}(\mathcal{P},\mathcal{L})| \le |\mathcal{P}|\,|\mathcal{L}|^{1/2} + |\mathcal{L}|\,.$$

Likewise, by doing the same proof with $\mathcal{P}$ and $\mathcal{L}$ switched (or by duality — we can switch points to lines and lines to points), we get that

$$|\mathcal{I}(\mathcal{P},\mathcal{L})| \le |\mathcal{P}|^{1/2}\,|\mathcal{L}| + |\mathcal{P}|\,.$$

In particular, this tells us that n points and n lines have $O(n^{3/2})$ incidences.

This relates to something we've seen before — our discussion about the extremal number of C_4's. (The fact that every pair of points determines at most one line corresponds to the bipartite graph on points and lines having no C_4.) There we saw that $\mathrm{ex}(n,C_4) = \Theta(n^{3/2})$, and the lower bound came from points and lines in $\mathbb{F}_p^2$. All the points and all the lines in $\mathbb{F}_p^2$ gave us this bound. But it was really important we worked in $\mathbb{F}_p$ and not $\mathbb{R}$, because this bound is *not* tight in Euclidean space — we'll now prove a better bound.

In order to prove a better bound, we'll have to go beyond the fact that every pair of points determines one line — that fact is true for finite fields as well. But in Euclidean space we also have topology of $\mathbb{R}^2$. We used this in proving the crossing number inequality (in particular, we used the fact that $|V|-|E|+|F|=2$, which doesn't make any sense in finite fields). And we can use this to get a better bound here.

Theorem 25.13 (Szemerédi–Trotter theorem)
For any sets $\mathcal{P}$ of points and $\mathcal{L}$ of lines in $\mathbb{R}^2$, we have

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

It turns out that this bound, up to a constant factor, is tight for *every* possible choice of a number of points and lines. Notably, when $|\mathcal{P}| = |\mathcal{L}| = n$, the upper bound is $|\mathcal{I}(\mathcal{P}, \mathcal{L})| = O(n^{4/3})$, which is better than the bound of $n^{3/2}$ we saw using essentially graph theory alone.

Here's an example of a tight lower bound.

Example 25.14

Take a grid of points $[k] \times [2k^2]$, and take the lines $y = mx + b$ where $m \in [k]$ and $b \in [k^2]$. There are $2k^3$ points and k^3 lines, and every line contains exactly k points, so $\mathcal{I}(\mathcal{P}, \mathcal{L}) = k^4$; this shows the bound is tight.

We have a couple of things to do. We'll prove the Szemerédi–Trotter theorem, and then use it to prove Elekes's bound for the sum-product problem.

Proof of Szemerédi–Trotter. First we'll do a bit of cleanup — remove all lines in $\mathcal{L}$ with at most one point of $\mathcal{P}$ (these lines are boring, and we'll add them back in the end).

Given a drawing of points and lines, we can turn this drawing into a graph in the visually obvious way — we keep the points of $\mathcal{P}$ as vertices of our graph, and we put in edges that are segments in this drawing. (We're ignoring the rays that protrude out; we keep our drawing, and view it as a graph.)

So now we have a graph $G = (V, E)$, where $V = \mathcal{P}$. Let's assume for now that $\mathcal{I}(\mathcal{P}, \mathcal{L}) \geq 8|\mathcal{P}|$ (because otherwise we already know trivially that the result is true). Then each line with k incidences has $k - 1 \geq \frac{k}{2}$ edges, so

$$|E| \geq \frac{|\mathcal{I}(\mathcal{P}, \mathcal{L})|}{2} \geq 4|V|.$$

Then by the crossing number inequality, we have that

$$\operatorname{cr}(G) \gtrsim \frac{|E|^3}{|V|^2} \gtrsim \frac{|\mathcal{I}(\mathcal{P}, \mathcal{L})|^3}{|\mathcal{P}|^2}.$$

On the other hand, what are the crossings in our drawing? The crossings come from pairs of lines, so the number of crossings is at most the number of pairs of lines; this means $\operatorname{cr}(G) \leq |\mathcal{L}|^2$. Combining these two inequalities together, we have that

$$|\mathcal{I}(\mathcal{P}, \mathcal{L})| \lesssim |P|^{2/3} |\mathcal{L}|^{2/3}.$$

We have to remember to add back in the easier linear contributions (from our assumption $|\mathcal{I}(\mathcal{P}, \mathcal{L})| \leq 8|\mathcal{P}|$ and from removing lines with at most one point) to obtain the full form. □

§25.2.3 Proof of sum-product estimate

We're now ready to prove the sum-product estimate stated earlier.

Proof of Elekes. The proof will be done by setting up a point-line incidence graph and applying the Szemerédi–Trotter theorem. We take

$$\mathcal{P} = \{(x, y) \mid x \in A + A, y \in A \cdot A\}$$

(essentially, we take the Cartesian product of the sumset and product set), and

$$\mathcal{L} = \{y = a(x - a') \mid a, a' \in A\}.$$

So we have a set with $A + A$ on the x-coordinates and $A \cdot A$ on the y-coordinates, and a bunch of lines given by the above equations.

For every line $y = a(x - a')$ in $\mathcal{L}$, let's think about what points it has in this set. This line is tailor-made so that we can put in some sumset elements as x and get back product set elements as y — the line is designed to go through lots of points in this grid. In particular, we see that every line has at least $|A|$ incidences, as we can set $x = a' + b$ and $y = ab$ for any $b \in A$.

So every line has lots of incidences, and then we can apply Szemerédi–Trotter — we have $|\mathcal{L}| = |A|^2$, so

$$|A|^3 = |A|\,|\mathcal{L}| \le |\mathcal{I}(\mathcal{P}, \mathcal{L})| \lesssim |\mathcal{P}|^{2/3}\,|\mathcal{L}|^{2/3} + |\mathcal{P}| + |\mathcal{L}| \lesssim |A + A|^{2/3}\,|A \cdot A|^{2/3}\,|A|^{4/3}$$

(one can check that the linear terms are less significant). Rearranging this gives the desired bound. □

§25.3 Solymosi's sum-product bound

In the second half of the lecture, we'll see an improved bound.

Theorem 25.15 (Solymosi)
Every $A \subseteq \mathbb{R}_{>0}$ satisfies

$$|A \cdot A|\,|A + A| \gtrsim \frac{|A|^4}{\log |A|}.$$

As a corollary, we have

$$\max\{|A + A|, |A \cdot A|\} \gtrsim |A|^{4/3 - o(1)}.$$

(This is a small improvement on the constant of $\frac{5}{4}$ we proved earlier.)

Remark 25.16. The bound of Solymosi (on $|A + A|\,|A \cdot A|$) is essentially tight — if we take $A = [n]$, then $|A + A| \approx n$ and $|A \cdot A|$ is slightly less than n^2. So the exponent of 4 cannot be improved.

The proof will use a concept called multiplicative energy, which is basically a version of the additive energy that we saw last time.

Definition 25.17. The *multiplicative energy* of A is defined as

$$E_\times(A) = \#\{(a, b, c, d) \in A^4 \mid ab = cd\}.$$

As with the additive energy, we can bound multiplicative energy in terms of the product set — we have

$$E_\times(A) = \sum_{x \in AA} \#\{(a, b) \in A^2 \mid ab = x\}^2 \ge \frac{|A|^4}{|AA|}$$

by Cauchy–Schwarz.

We define the *quotient set* as

$$A/A = \left\{\frac{a}{b} \mid a, b \in A\right\},$$

and we write $r(s) = \#\{(a, b) \in A \times A \mid s = a/b\}$. It's not hard to see that

$$E_\times(A) = \sum_{s \in A/A} r(s)^2.$$

The intuition is that it's easier to work with a situation where all the $r(s)$ are roughly the same size (as it's easier to work with regular graphs). There's a standard trick to get there — we just focus on a chunk of

values, and only think of that chunk (a subset of summands where all the r's are roughly the same). This method is called *dyadic pigeohnoling*.

By pigeonhole with a dyadic partition, there exists some nonnegative integer $k \lesssim \log |A|$ such that setting $D = \{s \mid 2^k \le r(s) \le 2^{k+1}\}$ and $m = |D|$, we have

$$\sum_{s \in D} r(s)^2 \gtrsim \frac{E_\times(A)}{\log |A|}.$$

(Essentially, we want to make sure that all our summands are roughly the same; we restrict to a dyadic interval and take the one with the biggest contribution; the factor of $\log |A|$ is from the number of possibilities for k.) On the other hand, because each summand is upper-bounded by 2^{k+1} and there are m summands, we have

$$\frac{E_\times(A)}{\log |A|} \lesssim \sum_{s \in D} r(s)^2 \le m 2^{2k+2}.$$

Now let the elements of D be $s_1 < \cdots < s_m$.

Now imagine the grid $A \times A$ (where we take the elements of A, and form a square grid of points). For each i, consider the line $\ell_i : y = s_i x$ — so we have a bunch of lines given by some slopes s_1, s_2, ... We should think of these lines as going through many points of A, since $r(s)$ (the number of points on this line) is pretty large (around 2^k).

Furthermore, we can draw one vertical line $\ell_{m+1} \colon x = \min A$.

Let $L_j = (A \times A) \cap \ell_j$ be the set of grid points that the corresponding line intersects — so we have a bunch of points along each line. We have $|L_j| = r(s_j) \ge 2^k$. Furthermore, ℓ_{m+1} is a vertical line, and every grid point on the previous line is also going to appear on the leftmost line, so $|L_{m+1}| \ge |L_m| \ge 2^k$.

Now let's think about what's going on in one of these sectors. As we add up points on one of the line and all the points on the next, we get some grid happening — the sumset of these two sets of points is a grid. And the number of points in this grid is the product of the numbers of points on the two lines — in other words,

$$|L_{j+1} + L_j| = |L_j| \, |L_{j+1}| \, .$$

Furthermore, if we look at what's happening in one sector, and then the next sector, and so on, they're all happening on disjoint parts of the plane. So these sets $L_{j+1} + L_j$ are all disjoint for distinct j — because they lie in different sectors, so geometrically they have nothing to do with each other. (They cannot overlap.)

Now we consider $|A + A|^2$ — we look at sumsets of points with one on ℓ_1 and one on ℓ_2, and so on. For each of these points, the coordinates of both of these points are in the sumset of A. So

$$|A + A|^2 = |(A \times A) + (A \times A)|$$

(where $A \times A$ denotes the cartesian product, and then we take the sumset of these two cartesian products). But we can just look at the contributions from each of the secotrs, so we get

$$|A + A|^2 = |(A \times A) + (A \times A)| \ge \sum_{j=1}^{m} |L_{j+1} + L_j| = \sum_{j=1}^{m} |L_j| \, |L_{j+1}| \ge m 2^{2k}$$

(since the sets $L_{j+1} + L_j$ are all disjoint, and we know $|L_j| \ge 2^k$). But we had an upper bound on the multiplicative energy based on 2^{2k}, so we get

$$|A + A|^2 \ge m 2^{2k} \gtrsim \frac{E_\times(A)}{\log |A|} \ge \frac{|A|^4}{|AA| \log |A|}.$$

And rearranging gives the desired bound.

§25.4 Historical remarks

Solymosi proved this bound in 2009, and for a long time this was the best bound for the sum-product problem (i.e., the best exponent was $\frac{4}{3}$). There was a breakthrough in 2015 where this bound was improved — Konyagin and Shkredov (2015) proved that

$$\max\{|A+A|, |AA|\} \geq |A|^{4/3+c}$$

for some small constant $c > 0$, and the small constant was improved several times. The best result may have $c \approx \frac{1}{2000}$; it's a small, but nonetheless positive, constant.

The methods are much more technical and intricate, but it's already interesting that despite all these new ideas, we are only able to improve on what we've seen by a tiny bit.

We're still very far away from the sum-product conjecture, which says that this exponent can be basically 2; that is a goal that for now seems very far away.

§25.5 Sum-product problem in $\mathbb{F}_p$

Finally, we'll close by mentioning a related but also important problem about what happens in finite fields. This started with an important work by Bourgain, Katz, and Tao.

Nothing we said in this lecture would work, since we relied heavily on the topology of $\mathbb{R}^2$, but other methods are relevant.

Theorem 25.18

For every $\varepsilon > 0$, there exist $\delta > 0$ and $c > 0$ such that for all $A \subseteq \mathbb{F}_p$ (with p prime) with $1 < |A| < p^{1-\varepsilon}$, we have

$$\max\{|A+A|, |AA|\} \geq c|A|^{1+\delta}.$$

We need a bit of room at the end — otherwise if $A = \mathbb{F}_p$ is the whole space, there is no room to grow. But this states that as long as there is room to grow, one of the sumset and product sets has to grow.

In the finite field setting, there's a relationship between $|A|$ and p, so additional interesting things happen. But it is important that p is prime — if p is instead a prime power, then this statement is actually false.

Question 25.19. What if $A \subseteq \mathbb{F}_{p^2}$ — is there a nontrivial set (not a singleton or the whole field) for which the sumset and product set do not expand?

We can take A to be the subfield $\mathbb{F}_p$ (which is contained in $\mathbb{F}_{p^2}$) — this is closed under multiplication and addition, so it doesn't expand. So it's very important that p is prime.

Qualitatively, what this says is that not only does the field $\mathbb{F}_p$ not have a subfield, but it also doesn't even have an *approximate* subfield — something that quantitatively looks closed in both addition and multiplication.

These types of results are important not just in combinatorics but also in analysis. Something similar is true over $\mathbb{R}$ — $\mathbb{C}$ has a subfield (namely $\mathbb{R}$), but $\mathbb{R}$ doesn't have a large subfield (i.e., one of positive dimension — of course there are $\mathbb{Q}$ and similar things). That fact is important in analysis, and was recently used in several major breakthroughs in problems in analysis. You can see that as an analytic analog of the kinds of problems we've seen today.

Next class we'll talk about the Green–Tao theorem, based on work in Prof. Zhao's PhD thesis. We'll see an overview of the main strategies used to prove the Green–Tao theorem — that the prime numbers contain arbitrarily long arithmetic progressions.

§26 December 13, 2023

§26.1 The Green–Tao theorem

Today we'll discuss the Green–Tao theorem.

Theorem 26.1 (Green–Tao)
The primes contain arbitrarily long arithmetic progressions.

This theorem was proved in 2004 and published 4 years later. It's a famous result, and one of the celebrated achievements of the century so far.

Prof. Zhao first heard about this result when he was a high school student; it sounds really cool and really deep. Indeed it is a deep result with lots of beautiful ideas. Prof. Zhao will tell us the core combinatorial idea behind the proof; we can't discuss the whole proofs (there are a lot of technical details, especially involving number theory), but we'll see an overview of the idea.

This is something Prof. Zhao thought about as a PhD student — when he started a PhD, he started thinking about problems related to sparse graph regularity. It was somewhat surprising that this work eventually led him to look at the proof of the Green–Tao theorem, and eventually come up with a simplification from the perspective of graph theory — using graph theoretic ideas, he could simplify some of the core combinatorial arguments. That's the main thing we'll see today — the overview of the proof, especially from this angle.

This is a theorem about prime numbers, and it's a result in number theory; certainly there are important number theoretic ideas. But a lot of the innovation is in combinatorics. They state in their abstract that the main new ingredient is a certain *transference principle*; we'll see this word later on. It allows us to deduce from Szemerédi's theorem that any subset of a sufficiently pseudorandom set of positive relative density contains progressions of arbitrary length. So that's the main combinatorial result of this paper; we'll state a precise version soon.

This looks like an extension of Szemerédi's theorem, which says that in $\mathbb{N}$, any positive density subset has long APs. This is about a sparse setting, where we start with a sufficiently pseudorandom sparse set and then take a subset of that; and we still have the Szemerédi property. (Green writes that the main advance lies not in our understanding of the primes, but rather what we can say about arithmetic progressions; and we'll focus on the innovations regarding APs, which has been a constant topic throughout this class.)

§26.2 Overview of proof strategy

First let's see a high-level overview of the proof strategy.

First, as a reminder, here's Szemerédi's theorem.

Theorem 26.2 (Szemerédi)
For every fixed $k \geq 3$, every k-AP-free subset of $[N]$ has size $o(N)$.

This is a theorem about dense subsets of the integers. On the other hand,

$$\#\{\text{primes} \leq N\} \sim \frac{N}{\log N}$$

by the prime number theorem. So at face value, Szemerédi's theorem doesn't say anything about arithmetic progressions in the primes. It's possible that just having better quantitative bounds on Szemerédi's theorem

might prove Green–Tao alone — we now have good enough bounds for 3-APs, but we still don't know whether density alone is enough for larger values of k.

Instead, the approach is the following. We're going to embed the primes into a slightly larger set, which we'll call the *almost primes* — roughly speaking, these are numbers without small prime divisors. For various reasons, these numbers are much easier to work with than the primes themselves. In the world of analytic number theory, these numbers are 'smoother' than the primes — if you think of the primes as being very discrete, this is almost like taking some convolution of the primes, giving you a smoother function that's much nicer to work with in the analytic number theory world.

And using pretty important and sophisticated ideas from number theory — in particular, developments related to small gaps between primes, which later lead to breakthroughs by Yitang Zhang on bounded gaps — they could analyze almost-primes, and understand various statistics well. For example, one could determine the density of k-APs in the almost primes, even though the density of k-APs in primes was unknown.

The almost primes are some set that one constructs; and it has a few nice properties.

(1) The set of primes has positive relative density in the almost primes —

$$\frac{\#\{\text{primes} \le N\}}{\#\{\text{almost primes} \le N\}} \ge \delta_k > 0.$$

(2) The almost primes behave pseudorandomly in a specific sense — with respect to certain linear pattern counts.

(We should think of k as fixed throughout; all these constructions, including the set of almost primes, depend on k.)

So this is the setup — instead of studying the primes, we embed them in a slightly larger set that's easier to analyze.

And then Green–Tao prove a *relative Szemerédi theorem*. (Throughout, think of $k \ge 3$ as fixed.) Informally, think of $S \subseteq \mathbb{Z}/N\mathbb{Z}$ as a fairly sparse set.

Theorem 26.3 (Relative Szemerédi theorem)

If S satisfies certain pseudorandomness conditions, then every k-AP-free subset of S has size $o(|S|)$.

Szemerédi's theorem is equivalent to this statement when $S = \mathbb{Z}/N\mathbb{Z}$ is the whole cyclic group; the relative Szemerédi theorem is about what happens when we start with much sparser but pseudorandom S.

We should think of S as the set of almost primes; then this allows us to deduce Szemerédi's theorem for the primes.

In particular, the Green–Tao theorem, following this strategy, tells us not only that the primes contain arbitrarily long APs, but even that any positive-density subset of the primes contain arbitrarily long APs (e.g. the primes that are 1 mod 100).

§26.3 Debiasing the primes

The primes are not exactly pseudorandom, since they have certain biases — for example, there's only one even prime, and mod 6 they can only be 1 or 5 (with finitely many exceptions). This makes the primes look a bit different than a random subset of integers. There's a standard method to de-bias them. The bias is mod small primes — for example, the parity of the primes is biased towards the odd numbers. And we can de-bias using the W-trick .

Let $w = w(N)$ be some number going to ∞ very slowly (e.g. $\log\log\log N$), and let $W = \prod_{p \le w} p$ be the product of all primes up to w. The W-trick is that instead of considering the set of primes, we consider the *W-tricked* set of primes:

Definition 26.4. The set of *W-tricked primes* is $\{n \mid nW + 1 \text{ prime}\}$.

Now this set no longer suffers from a parity bias — the fraction of n that are even is $\frac{1}{2}$, and the fraction of n that are 1 mod 3 is $\frac{1}{3}$, and so on. So this trick allows us to get rid of biases mod small numbers.

For example, if you just want to get rid of the mod 2 bias, instead of looking at integers you can just look at all odd integers. And if you want to get rid of the mod 3 bias, you can just look at all 1 mod 6 integers. And so on.

§26.4 A precise relative Roth theorem

We have a relative Szemerédi theorem; for simplicity we'll focus on a relative Roth theorem, the $k = 3$ case of the relative Szemerédi theorem.

We would like to state a precise version of the relative Roth theorem.

When we proved Roth's theorem using graph theory, we made a graph where the triangles in this graph corresponded to 3-APs. In that construction we had three sets X, Y, and Z, which were all $\mathbb{Z}/N\mathbb{Z}$ — so we had N vertices in each. And we built a graph, which we'll call G_S (where $S \subseteq \mathbb{Z}/N\mathbb{Z}$), where edges are given as follows:

- $x \sim y$ if and only if $2x + y \in S$;
- $x \sim z$ if and only if $x - z \in S$;
- $y \sim z$ if and only if $-y - 2z \in S$.

These expressions are chosen because three vertices $(x, y, z) \in X \times Y \times Z$ form a triangle if and only if $2x + y$, $x - z$, and $-y - 2z$ are all elements of S, and these three expressions form a 3-AP.

So we've defined a graph. The triangle density in this tripartite graph is what we'd expect (we respect the tripartite structure when we talk about densities — if we choose one vertex $x \in X$, $y \in Y$, and $z \in Z$, we look at the probability they form a triangle).

Here's our pseudorandomness condition.

Definition 26.5. We say $S \subseteq \mathbb{Z}/\mathbb{Z}$ satisfies the 3-*linear forms condition* (3LFC) with tolerance ε if for all $F \subseteq K_{2,2,2}$, we have

$$(1 - \varepsilon)p^{e(F)} \le t(F, G_S) \le (1 + \varepsilon)p^{e(F)}.$$

Here $t(F, G_S)$ is the density of F in G_S *respecting tripartiteness* — for example, if $F = K_{2,2,2}$, then we imagine choosing two vertices in X, two in Y, and two in Z and consider the probability that they form $K_{2,2,2}$.

If S were a random set with density p, you'd expect this to be roughly $p^{e(F)}$; and the condition states that this is correct up to a small error.

You should think of this as analogous to the C_4-density in Chung–Graham–Wilson ($C_4 = K_{2,2,2}$ is the 2-blowup of an edge, whereas here we look at the 2-blowup of a triangle).

Notably p here isn't necessarily constant — we should think of it as decaying with N (it's supposed to be something like the density of primes).

With this, we can formulate a precise version of the relative Roth theorem.

Theorem 26.6 (Relative Roth theorem, Conlon–Fox–Zhao)

For every $\delta > 0$, there exists $\varepsilon > 0$ and N_0 such that for all odd $N \geq N_0$, if $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies 3LFC with tolerance ε, then every 3-AP-free subset of S has size at most $\delta |S|$.

This is a precise version of our more informal statement from earlier, specifically for 3-APs.

The original Green–Tao paper proved a theorem that looks like this, but they required not only the 3LFC but also a more technical correlation condition (and part of Prof. Zhao's PhD thesis was getting rid of that; part of the insight was that viewing things as a graph was very helpful).

Remark 26.7. What about k-APs instead of 3-APs? You can generalize the statement almost verbatim. You have to come up with the correct definition of the kLFC, but that's informed by what we've seen already — to prove Szemerédi's theorem for 4-APs, you build a 4-partite 3-uniform hypergraph. And we can write down a similar statement where instead of the blowup of a triangle, we look at the blowup of a simplex.

Proving Szemerédi's theorem is much more difficult for 4-APs compared to 3-APs, and even more difficult for longer APs. But for relative Szemerédi's theorem, there actually isn't a substantial difficulty difference for different values of k. (We will assume Szemerédi's theorem as a black box.) Still, we'll focus on 3-APs for notational simplicity.

Remark 26.8. Quite surprisingly, even though this is an extension of Szemerédi's theorem to sparser sets, whereas Szemerédi's theorem is about dense sets, in the proof method we'll be able to assume Szemerédi's theorem as a black box, without opening up its proof; and we'll be able to transfer that result from the dense setting to the sparse setting.

This doesn't sound believable — usually if you prove something in the dense setting, in the sparse setting you have to open up the proof and do things over again. But here we'll have some techniques — the transference principle — that allow us to bring down a dense setting result to a sparse setting result. This is quite magical.

§26.5 A digression — random sets

There is a separate and interesting problem (not directly related to the rest of the discussion on Green–Tao) about what happens to Szemerédi's theorem in a *random* set — instead of taking S with certain pseudorandomness conditions, what if we just take a random S? This was open for quite some time; now we understand the situation almost completely. Here's the theorem.

Theorem 26.9 (Conlon–Gowers 2016, Schacht 2016)

For every $k \geq 3$ and $\delta > 0$, there exists C such that if $p > CN^{-1/(k-1)}$, then with probability $1 - o(1)$ (as $N \to \infty$), given a random $S \subseteq [N]$ where each element is included with probability p, every k-AP-free subset of S has size at most $\delta |S|$.

It turns out this exponent of $-1/(k-1)$ is best possible.

If you didn't know this result and tried to deduce something like this from the relative Szemerédi theorem, you can get something, but you'll have a worse exponent. If you take a random subset of $[N]$ (or $\mathbb{Z}/N\mathbb{Z}$), then by a second moment calculation, you will have the 3LFC condition if p is not too small. You won't get this exponent, but you will get *some* constant exponent.

Remark 26.10. The threshold $CN^{-1/(k-1)}$ is the best possible (up to the constant C) — this is because the expected number of k-APs in S is $O(p^k N^2)$ by linearity of expectation. ANd if $p < cN^{-1/(k-1)}$ (where c is a small constant), then this quantity becomes less than $\frac{1}{2}\mathbb{E}\,|S| = \frac{1}{2}Np$. That means you can start with S and then just get rid of all its k-APs, while still keeping more than half its elements; that violates the conclusion. (This is the deletion method we saw when constructing H-free graphs via the probabilistic method in the second chapter.)

Remark 26.11. This theorem and the techniques that went into it are related to the *hypergraph container method*, a modern method developed only about 10 years ago that could also be used to prove the same result. This method itself has had a big impact in extremal combinatorics; recently the authors who proved it won the Steele prize, a big honor.

Remark 26.12. How does the 3LFC condition relate to other definitions of pseudorandomness of sparse graphs?

One problem with sparse graphs is that counting is hard. Here we're trying to count triangles; discrepancy type conditions aren't good enough for that. But it would help to have something even better than counting triangles. This is like a second moment condition about counting triangles, which gives us more room to count triangles.

Remark 26.13. How hard is the $K_{2,2,2}$ condition to check for the almost primes? This is a major part of Green–Tao's innovation — they realized it's much easier to verify these conditions in the almost primes. Prof. Zhao doesn't have too much intuition about the analytic number theory, but the intuition is that the almost primes are smoother than the primes.

It turns out that we now do know, because of follow-up work by Green–Tao involving much more difficult mathematics, that the primes *also* satisfy the 3LFC condition — so the number of patterns in the primes really is asymptotically what you'd get if you treat the primes as random. (With one caveat — we don't know if things like Goldbach or Twin Prime are true.)

§26.6 An outline

The goal of the next part of the lecture is to sketch a proof fo the relative Roth theorem. We'll explain the beautiful idea of the transference principle, which allows us to take Szemerédi's theorem — about dense subsets of integers — and transfer it as a black box to sparse sets.

We're given $A \subseteq S$ where A has positive relative density — $|A| \geq \delta\,|S|$ — and here S satisfies the 3LFC. (Here S is a subset of $\mathbb{Z}/N\mathbb{Z}$.)

The idea is as follows.

(1) We'll approximate A by a dense model — this means there will be a dense set $B \subseteq \mathbb{Z}/N\mathbb{Z}$ such that

$$\frac{|B|}{N} \approx \frac{|A|}{|S|} \geq \delta.$$

This set B should have some additional properties — B should be 'close' to A in some sense, that will be related to the 'cut norm' — we'll define this precisely for integers, but it's very much related to the cut norm we saw earlier in the chapter on graph limits.

(2) Then we'll count k-APs. The key part here is a sparse counting lemma, which will show that the number of k-APs in A is approximately equal to the number of k-APs in B, normalized appropriately

— i.e., we have

$$\#\{k\text{-APs in } A\} \approx p^k \#\{k\text{-APs in } B\}$$

where $p = |S|/N$. (This is what you'd expect if A and B are good models of each other.)

(3) Now, B is a dense subset of $[N]$, and Szemerédi's theorem as a black box tells us that the number of k-APs in B is pretty large. (Szemerédi's theorem really tells us that B contains at least one k-AP. But we saw a homework problem that shows not only does B contain at least one k-AP, but it must contain quadratically many — this was a supersaturation argument.) Then the number of k-APs in A is also quite large (at least $p^k N^2$, up to a constant depending on δ), so in particular it is greater than 0. (There's a small thing we have to check in that this counts trivial k-APs as well, but there are very few of those.)

We'll now explain more precisely what the statements are that we're proving. There's two statements we need — one is coming up with a dense model, and the other is a sparse counting lemma.

§26.7 The dense model theorem

We're familiar with the cut norm for a graphon — we look at one set on the x-axis and one on the y-axis, and consider the integral on that box.

Suppose we're given a function $f: \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$. Here we want to define a cut norm, but we only have one dimension; the idea is to consider the Cayley graph, and look at the cut norm of this graph.

Definition 26.14. We define $\|f\| = \sup_{A,B \subseteq \mathbb{Z}/N\mathbb{Z}} |\mathbb{E}_{x,y \in \mathbb{Z}/N\mathbb{Z}} f(x+y)\mathbf{1}_A(x)\mathbf{1}_B(y)|$.

In other words, this is the cut norm of the graphon $(x, y) \mapsto f(x + y)$.

Theorem 26.15 (Dense model theorem)

For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds: for every $A \subseteq S \subseteq \mathbb{Z}/N\mathbb{Z}$, letting $p = |S|/N$, if $\|\mathbf{1}_S - p\|_\square \le \delta p$, then there exists $g: \mathbb{Z}/N\mathbb{Z} \to [0, 1]$ such that

$$\|\mathbf{1}_A - pg\|_\square \le \varepsilon p.$$

Think of p as some density (decaying with N). This says that if $\mathbf{1}_S$ is close to the constant p in cut norm, then there exists some g such that A can be approximated by pg in cut norm.

Think of g as a weighted version of B — B is required to be a set, but there's not a big difference between a set and a function taking values in $[0, 1]$ (you can go from one to the other by taking an indicator function or by random sampling).

But the point is that we start with S, which itself could be quite sparse. But we assume that S is close to a constant — so S is pseudorandom in some discrepancy-like sense (this is actually much weaker than the 3LFC condition, and it follows from our hypothesis). And then the conclusion is we can find a dense model for A. Here A is an arbitrary subset of S, so A itself is not pseudorandom. But also, A is sparse — think of S as being tiny, and then A is some very sparse subset of S. And we want to approximate A by a dense set — we're going to find a dense subset B that approximates A in cut norm. (Think of g as the indicator function of some set B.)

Remark 26.16. Intuitively, why do we take $f(x + y)$ in the definition of the cut norm for f?

In the graph that we drew, we're going to talk about the discrepancy between two parts; and between these parts we'll have some graph constructed via some formula, that looks roughly like $x + y$. (We had coefficients, but the coefficients don't matter much.) The condition on cut norm really says that the graph satisfies a discrepancy condition.

Remark 26.17. We should think of S as a sparse pseudorandom subset of S, and A as an arbitrary relatively dense subset of S; and this says we can model A by an actually dense set.

So that's the statement of the dense model theorem. We won't prove it; the proof is not very long, but uses some clever ideas. In particular, it uses the hyperplane separation theorem and the Weirstrass polynomial approximation theorem; it's neat, but we won't have time to do it in class.

§26.8 The sparse counting lemma

We'll now jump to the next part, the sparse counting lemma.

This really is the crux of a lot of the difficulties. When we were discussing quasirandom graphs earlier in the semester, we mentioned the quasirandom graphs theorem is really about dense graphs; many of the implications there break down for sparse graphs, in particular the counting lemma. So the sparse counting lemma is a way to remedy those technical difficulties, but with additional assumptions — namely the assumption on S that we have here.

The setup is that we're going to have weighted tripartite graphs. We'll have three different weighted graphs f, g, and ν. Here ν corresponds to the set $\frac{1}{p}\mathbf{1}_S$ (we need to properly normalize so that we have a function of constant-order expectation). And g is going to be the function coming out of the dense model, which we should think of as $\mathbf{1}_B$ (the indicator function of our dense model), and f will be $\frac{1}{p}\mathbf{1}_A$ (the properly normalized indicator function for A).

We can then talk about subgraph densities for any of these functions. The setup will require several assumptions.

- $0 \le f \le \nu$ — this is because $A \subseteq S$. (Note that ν is not capped — it can take arbitrarily large values.)
- $0 \le g \le 1$ — this is because g is our dense model, which has no normalization.
- The 3LFC condition, which states that $|t(F, \nu) - 1| \le \varepsilon$ whenever $F \subseteq K_{2,2,2}$.
- $\|f - g\|_\square \le \varepsilon$. (This is the conclusion of the dense model theorem from earlier.)

Theorem 26.18 (Sparse triangle counting lemma, Conlon–Fox–Zhao)
In the above setup, we have $|t(K_3, f) - t(K_3, g)| \le \varepsilon^{\Omega(1)}$.

If we assume the sparse triangle counting lemma, we can complete our program — remember that there's a graph we set up where 3-APs correspond to triangles. Then having similar triangle counts tells us that the 3-AP count in A matches the one in B, properly normalized; and that finishes off the rest of the proof.

Now we'll discuss some of the ideas in the sparse triangle counting lemma. First, here's a review of how the triangle counting lemma worked in the dense setting (in the graphon chapter).

In the dense setting, where $\nu = 1$ (so f and g are both bounded between 0 and 1, and f and g are close in cut norm), the statement was that

$$|t(K_3, f) - t(K_3, g)| \le 3\|f - g\|_\square \le 3\varepsilon.$$

The proof went as follows. We start with trying to count triangles in f, which has three parts with f between all three parts. And then up to an ε-loss, this is the same as counting triangles where we replace one of the bipartite graphs with g. (This is because the difference in neighborhoods is something bounded by the cut norm.) And we can do this two more times to get g in all three of the bipartite graphs.

More specifically, we're using the fact that if we fix z, then

$$\int (f(xy) - g(xy))f(xz)f(yz) \le \|f - g\|_\square.$$

What goes wrong if we remove the dense setting hypothesis and allow ν to be arbitrary, which means f can now be unbounded? Now we have a normalization of $\frac{1}{p}$, so f is unbounded; and then the two factors $f(xz)$ and $f(yz)$ can be large. So we cannot apply the cut norm bound to this expression. And so the above inequality is problematic (and can be completely false).

That's the difficulty we'd like to overcome. At a high level, the way we overcome this is to note that we start with a problem where we're counting triangles in two different settings, but there's some sparse host graph ν underneath (corresponding to S). We're going to perform several calculations that one at a time get rid of one of these ν's, and converts them to 1's — we do the same process, but replacing one of the sparse hosts by a dense host. (It should not be clear what this step is about, but this is a high level overview for now.) And we can keep on doing this until the problem reduces down to a setting where all three of the hosts are dense, and now we're back to the dense setting, which we already know how to handle.

How exactly do we do this? This is where we're going to use the 3LFC condition a lot, with a lot of Cauchy–Schwarz.

Roughly, the idea is the following. We start with trying to count triangles in a tripartite graph, with f between all three parts. And the difference between counting triangles with f and counting triangles with g can be separated out into two steps.

First, write $f_\wedge(x,y) = \mathbb{E}_z f(xz)f(yz)$ — this is the codegree between x and y. Then this difference (of triangle counts for f and g) is

$$\langle f, f_\wedge\rangle - \langle g, g_\wedge\rangle = \langle f, f_\wedge - g_\wedge\rangle - \langle f-g, g_\wedge\rangle.$$

The second term is easier to handle, because g is bounded by 1 — so then this second term is at most ε in absolute value, by the same argument we had earlier.

The difficulty is the term $\langle f, f_\wedge - g_\wedge\rangle$. In order to deal with this, we can apply Cauchy–Schwarz — by Cauchy–Schwarz, we deduce that

$$|\langle f, f_\wedge - g_\wedge\rangle|^2 = \mathbb{E}[f(f_\wedge - g_\wedge)]^2 \le \mathbb{E}[f(f_\wedge - g_\wedge)^2]\mathbb{E}f.$$

And the square $(f_\wedge - g_\wedge)^2$ is always nonnegative, so now we can replace f by ν — so we get that this is at most

$$\mathbb{E}[\nu(f_\wedge - g_\wedge)^2]\mathbb{E}[\nu].$$

(This is the important step — we got to replace f by ν, which we couldn't have done previously, because now the other factor is always nonnegative.)

The significance of this step might not be clear yet, but we'll try to sketch why this is useful. We should think of ν as being close to 1 — it's a pseudorandom function, so it should be close to 1 in many reasonable ways. And in fact, via some more Cauchy–Schwarz calculations which we will not do here, you can prove that

$$\mathbb{E}[\nu(f_\wedge - g_\wedge)^2] \approx \mathbb{E}[(f_\wedge - g_\wedge)^2].$$

(You should think of ν as being very pseudorandom, and therefore similar to 1; and you can justify this using Cauchy–Schwarz.)

Now what's going on with $\mathbb{E}[(f_\wedge - g_\wedge)^2]$? We can expand this into several terms as

$$\langle f_\wedge, f_\wedge\rangle - \langle f_\wedge, g_\wedge\rangle - \langle g_\wedge, f_\wedge\rangle + \langle g_\wedge, g_\wedge\rangle.$$

Each of these terms represents some 4-cycle density (counting two paths in f and two paths in g, or some combination thereof). And we'd like to show that all of these terms are approximately equal to each other.

Let's focus on teh most interesting case, of the first term (this is most interesting because both our objects $f_\wedge$ are sparse, and the sparse objects are the ones that we have the most trouble dealing with.) So we're counting 4-cycles in the f-graph.

And the key observation is that we can view this picture as a picture where we have a tripartite graph, with f between two pairs, but $f_\wedge$ between the third — and then the problem reduces to triangle counting in this graph.

It seems we've just gone from triangle counting to triangle counting for a different graph. But this graph is much nicer — the original f comes from a sparse edge set, but $f_\wedge$ is actually smoothed out (it comes from codegrees), so it behaves like a dense function — in other words, $f_\wedge \le 1$ in most places. (In a sparse pseudorandom graphs, whether there's an edge betwen two vertices can vary depending on what vertices you take. But the codegree is much more predictable, because you're averaging over lots of other vertices. And that analytically translates to this graph being much smoother, and looking much more like a dense graph.)

And that's what we mean in the step where we replace ν with 1 — we've done calculations and transformed the problem where all three pairs are sparse, to another problem where one of the pairs now became dense. And that is progress! And we can do it again, and again, and again, and get all the way down to the dense case. And that finishes off the sparse triangle counting lemma.