Metric embeddings and geometric inequalities

Lectures by Professor Assaf Naor
Scribes: Holden Lee and Kiran Vodrahalli

April 27, 2016

Contents

Introduction

The topic of this course is geometric inequalities with applications to metric embeddings; and we are actually going to do things more general than metric embeddings: Metric geometry. Today I will roughly explain what I want to cover and hopefully start proving a first major theorem. The strategy for this course is to teach novel work. Sometimes topics will be covered in textbooks, but a lot of these things will be a few weeks old. There are also some things which may not have been even written yet. I want to give you a taste for what's going on in the field. Notes from the course I taught last spring are also available.

One of my main guidelines in choosing topics will be topics that have many accessible open questions. I will mention open questions as we go along. I'm going to really choose topics that have proofs which are entirely self-contained. I'm trying to assume nothing. My intention is to make it completely clear and there should be no steps you don't understand.

Now this is a huge area. I will explain some of the background today. I'm going to use proofs of some major theorems as excuses to see the lemmas that go into the proofs. Some of the theorems are very famous and major, and you're going to see some improvements, but along the way, we will see some lemmas which are *immensely powerful*. So we will always be proving a concrete theorem. But actually somewhere along the way, there are lemmas which have wide applicability to many many areas. These are excuses to discuss methods, though the theorems are important.

The course can go in many directions: If some of you have some interests, we can always change the direction of the course, so express your interests as we go along.

Chapter 1

The Ribe Program

2/1/16

1 The Ribe Program

The main motivation for most of what we will discuss is called the Ribe Program, which is a research program many hundreds of papers large. We will see some snapshots of it, and it all comes from a theorem from 1975, **Ribe's rigidity theorem** 1.1.9, which we will state now and prove later in a modern way. This theorem was Martin Ribe's dissertation, which started a whole direction of mathematics, but after he wrote his dissertation he left mathematics. He's apparently a government official in Sweden. The theorem is in the context of Banach spaces; a relation between their linear structure and their structure as metric spaces. Now for some terminology.

Definition 1.1.1 (Banach space)**:** A **Banach space** is a complete, normed vector space. Therefore, a Banach space is equipped with a metric which defines vector length and distances between vectors. It is complete, so every Cauchy sequence of converges to a limit defined inside the space.

Definition 1.1.2: Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ be Banach spaces. We say that X is (crudely) **finitely representable** in Y if there exists some constant $K > 0$ such that for every finite-dimensional linear subspace $F \subseteq X$, there is a linear operator $S : F \to Y$ such that for every $x \in F$,

$$\|x\|_X \le \|Sx\|_Y \le K \|x\|_X .$$

Note K is decided once and for all, before the subspace F is chosen.

(Some authors use "finitely representable" to mean that this is true for any $K = 1 + \varepsilon$. We will not follow this terminology.)

Finite representability is important because it allows us to conclude that X has the same finite dimensional linear properties (**local properties**) as Y. That is, it preserves any invariant involves finitely many vectors, their lengths, etc.

Let's introduce some local properties like type. To motivate the definition, consider the triangle inequality, which says

$$\|y_1 + \cdots + y_n\|_Y \le \|y_1\|_Y + \cdots + \|y_n\|_Y .$$

In what sense can we improve the triangle inequality? In L^1 this is the best you can say. In many spaces there are ways to improve it if you think about it correctly.

For any choice $\varepsilon_1, \ldots, \varepsilon_n \in \{\pm 1\}$,

$$\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y \le \sum_{i=1}^n \|y_i\|_Y .$$

Definition 1.1.3: df:type Say that X has **type** p if there exists $T > 0$ such that for every $n, y_1, \ldots, y_n \in Y$,

$$\mathop{\mathbb{E}}_{\varepsilon\in\{\pm 1\}^n} \left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y \le T \left[\sum_{i=1}^n \|y_j\|_Y^p\right]^{\frac{1}{p}} .$$

The L^p norm is always at most the L^1 norm; if the lengths are spread out, this is asymptotically much better. Say Y has **nontrivial type** if $p > 1$.

For example, $L_p(\mu)$ has type $\min(p, 2)$.

Later we'll see a version of "type" for metric spaces. How far is the triangle inequality from being an equality is a common theme in many questions. In the case of normed spaces, this controls a lot of the geometry. Proving a result for $p > 1$ is hugely important.

Proposition 1.1.4: pr:finrep-type If X is finitely representable and Y has type p then also X has type p.

Proof. Let $x_1, \ldots, x_n \in X$. Let $F = \operatorname{span}\{x_1, \ldots, x_n\}$. Finite representability gives me $S : F \to Y$. Let $y_i = Sx_i$. What can we say about $\sum \varepsilon_i y_i$?

$$\begin{aligned}
\mathop{\mathbb{E}}_{\varepsilon} \left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y &= \mathop{\mathbb{E}}_{\varepsilon} \left\|S(\sum_{i=1}^n \varepsilon_i x_i)\right\|_Y \\
&\ge \mathop{\mathbb{E}}_{\varepsilon} \left\|\sum_{i=1}^n \varepsilon_i X_i\right\|_X \\
\mathop{\mathbb{E}}_{\varepsilon} \left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y &\le T \left(\sum_{i=1}^n \|Sx_i\|^p\right)^{\frac{1}{p}} \\
&\le TK \left(\sum_{i=1}^n \|x_i\|^p\right)^{\frac{1}{p}} .
\end{aligned}$$

Putting these two inequalities together gives the result. □

Theorem 1.1.5 (Kahane's inequality)**.** *For any normed space Y and $q \ge 1$, for all n, $y_1, \dots, y_n \in Y$,*

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i y_i\right\| \gtrsim_q \left(\mathbb{E}\left[\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y^q\right]\right)^{\frac{1}{q}}.$$

Here $\gtrsim_q$ means "up to a constant"; subscripts say what the constant depends on. The constant here does not depend on the norm Y.

Kahane's Theorem tells us that the LHS of Definition 1.1.3 can be replaced by any norm, if we change $\le$ to $\lesssim$. We get that having type p is equivalent to

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y^p \lesssim T^p \sum_{i=1}^n \|y_i\|_Y^p.$$

Recall the **parallelogram identity** in a Hilbert space H:

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|^2 = \sum_{i=1}^n \|y_i\|_H^2.$$

A different way to understand the inequality in the definition of "type" is: how far is a given norm from being an Euclidean norm? The **Jordan-von Neumann Theorem** says that if parallelogram identity holds then it's a Euclidean space. What happes if we turn it in an inequality?

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_H^2 \begin{matrix}\ge \\ \le\end{matrix} T \sum_{i=1}^n \|y_i\|_H^2.$$

Either inequality *still* characterizes a Euclidean space.

What happens if we add constants or change the power? We recover the definition for type and cotype (which has the inequality going the other way):

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_H^q \begin{matrix}\gtrsim \\ \lesssim\end{matrix} \sum_{i=1}^n \|y_i\|_H^q.$$

Definition 1.1.6: Say it has **cotype** q if

$$\sum_{i=1}^n \|y_i\|_Y^q \lesssim C^q \mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i y_i\right\|_Y^q$$

R. C. James invented the local theory of Banach spaces, the study of geometry that involves properties involving finitely many vectors ($\forall x_1, \dots, x_n, P(x_1, \dots, x_n)$ holds). As a counterexample, reflexivity cannot be characterized using finitely many vectors (this is a theorem).

Ribe discovered link between metric and linear spaces.

First, terminology.

Definition 1.1.7: Two Banach spaces are **uniformly homeomorphic** if there exists $f : X \to Y$ that is 1-1 and onto and f, f^{-1} are uniformly continuous.

Without the word "uniformly", if you think of the spaces as topological spaces, all of them are equivalent. Things become interesting when you quantify! "Uniformly" means you're controlling the quantity.

Theorem 1.1.8 (Kadec)**.** *Any two infinite-dimensional separable Banach spaces are homeomorphic.*

This is a amazing fact: these spaces are all topologically equivalent to Hilbert spaces!

Over time people people found more examples of Banach spaces that are homeomorphic but not uniformly homeomorphic. Ribe's rigidity theorem clarified a big chunk of what was happening.

Theorem 1.1.9 (Rigidity Theorem, Martin Ribe (1975))**.** thm:ribe *Suppose that X, Y are uniformly homeomorphic Banach spaces. Then X is finitely representable in Y and Y is finitely representable in X.*

For example, for L^p and L^q, for $p \neq q$ it's always that case that one is not finitely representable in the other, and hence by Ribe's Theorem, L^p, L^q are not uniformly homeomorphic. (When I write L_p, I mean $L_p(\mathbb{R})$.)

Theorem 1.1.10. *For every $p \geq 1, p \neq 2$, L_p and ℓ_p are finitely representable in each other, yet not uniformly homeomorphic.*

(Here ℓ_p is the sequence space.)

Exercise 1.1.11: Prove the first part of this theorem: L_p is finitely representable in ℓ_p.

Hint: approximate using step functions. You'll need to remember some measure theory.

When $p = 2$, L_p, ℓ_p are separable and isometric.

The theorem in various cases was proved by:

1. $p = 1$: Enflo
2. $1 < p < 2$: Bourgain
3. $p > 2$: Gorelik, applying the Brouwer fixed point theorem (topology)

Every linear property of a Banach signs which is local (type, cotype, etc.; involving summing, powers, etc.) is preserved under a general nonlinear deformation.

After Ribe's rigidity theorem, people wondered: can we reformulate the local theory of Banach spaces without mentioning anything about the linear structure? Ribe's rigidity theorem is more of an existence statement, we can't see anything about an explicit dictionary which maps statements about linear sums into statements about metric spaces. So people started to wonder whether we could reformulate the local theory of Banach spaces, but only looks at distances between pairs instead of summing things up. Local theory is one of the hugest subjects in analysis. If you could actually find a dictionary which takes one linear theorem at a time, and restate it with only distances, there is a huge potential here! Because the definition of type only involves distances between points, we can talk about a metric space's type or cotype. So maybe we can use the intution given by linear arguments, and then state things for metric spaces which often for very different reasons remain true from the linear domain. And then now maybe you can apply these arguments to graphs, or groups! We could be able to prove things about the much more general metric spaces. Thus, we end up applying theorems on linear spaces in situations with *a priori nothing* to do with linear spaces. This is massively powerful.

There are very crucial entries that are missing in the dictionary. We don't even now how to define many of the properties! This program has many interesting proofs. Some of the most interesting conjectures are how to define things!

Corollary 1.1.12. cor:uh-type *If X, Y are uniformly homeomorphic and if one of them is of type p, then the other does.*

This follows from Ribe's Theorem 1.1.9 and Proposition 1.1.4. Can we prove something like this theorem without using Ribe's Theorem 1.1.9? We want to reformulate the definition of type using only the distance, so this becomes self-evident.

Enflo had an amazing idea. Suppose X is a Banach space, $x_1, \dots, x_n \in X$. The type p inequality says

eq:type-p
$$\mathbb{E}\left[\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|^p\right] \lesssim_X \sum_{i=1}^n \|x_i\|^p . \tag{1.1}$$

Let's rewrite this in a silly way. Define $f : \{\pm 1\}^n \to X$ by

$$f(\varepsilon_1, \dots, \varepsilon_n) = \sum_{i=1}^n \varepsilon_i x_i.$$

Write $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$. Multiplying by 2^n, we can write the inequality (1.1) as

eq:type-gen
$$\mathbb{E}\left[\|f(\varepsilon) - f(-\varepsilon)\|^p\right] \lesssim_X \sum_{i=1}^n \mathbb{E}\left[\|f(\varepsilon) - f(\varepsilon_1, \dots, \varepsilon_{i-1}, -\varepsilon_i, \varepsilon_{i+1}, \dots, \varepsilon_n)\|^p\right]. \tag{1.2}$$

This inequality just involves distances between points $f(\varepsilon)$, so it is the reformulation we seek.

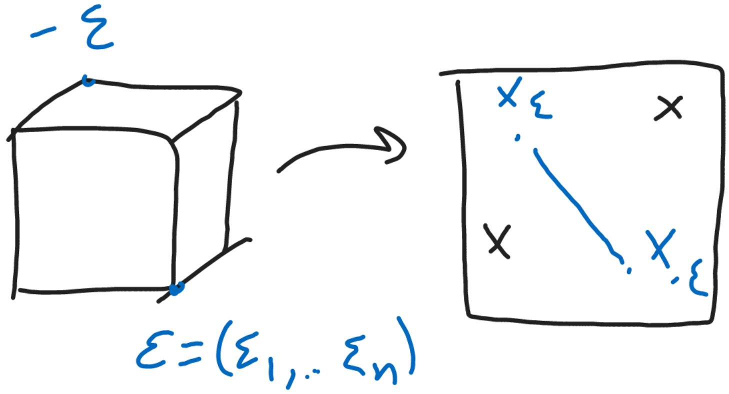
Definition 1.1.13: df:enflo A metric space (X, d_X) has **Enflo type** p if there exists $T > 0$ such that for every n and every $f : \{\pm 1\}^n \to X$,

$$\mathbb{E}[d_X(f(\varepsilon), f(-\varepsilon))^p] \le T^p \sum_{i=1}^n \mathbb{E}[d_X(f(\varepsilon), f(\varepsilon_1, \dots, \varepsilon_{i-1}, -\varepsilon_i, \varepsilon_{i+1}, \dots, \varepsilon_n))^p].$$

This is bold. It wasn't true before for a general function! The discrete cube (Boolean hypercube) $\{\pm 1\}^n$ is all the ϵ vectors, of which there are 2^n. Our function just assigns 2^n points arbitrarily. No structure whatsoever. As they are indexed this way, you see nothing. But you're free to label them by vertices of the cube however you want. But there are many labelings! In (1.2), the points had to be vertices of a cube, but in Definition 1.1.13, they are arbitrary. The moment you choose the labelings, you impose a cube structure between the points. Some of them are diagonals of the cube, some of them are edges. ϵ and $-\epsilon$ are antipodal points. But it's not really a diagonal. They are points on a manifold, and are a function of how you decided to label them. What this sum says is that the sum over all diagonals, the length of the diagonals to the power p is less than the sum over edges to the p^{th} powers (these are the points where one ϵ_i is different). Thus we can see

$$\sum \text{diag}^p \lesssim_X \sum \text{edge}^p.$$

The total pth power of lengths of diagonals is up to a constant, at most the same thing over all edges.



This is a vast generalization of type; we don't even know a Banach space satisfies this. The following is one of my favorite conjectures.

Conjecture 1.1.14 (Enflo)**.** *If a Banach space has type p then it also has Enflo type p.*

This has been open for 40 years. We will prove the following.

Theorem 1.1.15 (Bourgain-Milman-Wolfson, Pisier)**.** *If X is a Banach space of type $p > 1$ then X also has type $p - \varepsilon$ for every $\varepsilon > 0$.*

If you know the type inequality for parallelograms, you get it for arbitrary sets of points, up to ε. Basically, you're getting arbitrarily close to p instead of getting the exact result. We also know that the conjecture stated before is true for a lot of specific Banach spaces, though we do not yet have the general result. For instance, this is true for the L_4 norm. Index functions by vertices; some pairs are edges, some are diagonals; then the L^4 norm of the diagonals is at most that of the edges.

How do you go from knowing this for a linear function to deducing this for an arbitrary function?

Once you do this, you have a new entry in the hidden Ribe dictionary. If X and Y are uniformly homeomorphic Banach spaces and Y has Enflo type p, then so is X. The minute you throw away the linear structure, Corollary 1.1.12 becomes easy. It requires a tiny argument. Now you can take a completely arbitrary function $f : \{\pm 1\}^n \to X$. There exists a homeomorphism $\psi : X \to Y$ such that ψ, ψ^{-1} are uniformly continuous. Now we want to deduce that the same inequality in Y gives the same inequality in X.

Proposition 1.1.16: pr:uh-enflo If X, Y are uniformly homeomorphic Banach spaces and Y has Enflo type p, then so does X.

This is an example of making the abstract Ribe theorem explicit.

Lemma 1.1.17 (Corson-Klee)**.** *lem:corson-klee* *If X, Y are Banach spaces and $\psi : X \to Y$ are uniformly continuous, then for every $a > 0$ there exists $L(a)$ such that*

$$\|x_1 - x_2\|_X \geq a \implies \|\psi(x_1) - \psi(x_2)\| \leq L \|x_1 - x_2\|.$$

Proof sketch of 1.1.16 given Lemma 1.1.17. By definition of uniformly homeomorphic, there exists a homeomorphism $\psi : X \to Y$ such that ψ, ψ^{-1} are uniformly continuous. Lemma 1.1.17 tells us that ψ perserves distance up to a constant. Dividing so that the smallest nonzero distance you see is at least 1, we get the same inequality in the image and the preimage. $\square$

Proof details. Let ε^i denote ε with the ith coordinate flipped. We need to prove

$$\mathbb{E}(d_X(f(\varepsilon), f(-\varepsilon))^p) \leq T_f^p \sum_{i=1}^n \mathbb{E}(d_X(f(\varepsilon), f(\varepsilon^i))^p)$$

Without loss of generality, by scaling f we may assume that all the points $f(\varepsilon)$ are distance at least 1 apart. (X is a Banach space, so distance scales linearly; this doesn't affect whether the inequality holds.)

Let $\psi : X \to Y$ be such that ψ, ψ^{-1} are uniform homeomorphisms. Because ψ^{-1} is uniformly homeomorphic, there is C such that $d_Y(y_1, y_2) \leq 1$ implies $d_X(\psi^{-1}(y_1), \psi^{-1}(y_2)) < C$. WLOG, by scaling f we may assume that all the points $f(\varepsilon)$ are $\max(1, C)$ apart, so that the points $\psi \circ f(\varepsilon)$ are at least 1 apart.

We know that for any $g : \{\pm 1\}^n \to Y$ that

$$\mathbb{E}(d_X(g(\varepsilon), g(-\varepsilon))^p) \leq T_g^p \sum_{i=1}^n \mathbb{E}(d_X(g(\varepsilon), g(\varepsilon^i))^p).$$

We apply this to $g = \psi \circ f$,

$$\begin{tikzcd} & X \arrow[dd, "\psi"] \\ \{\pm 1\}^n \arrow[ur, "f"] \arrow[dr, "g=\psi\circ f"'] & \\ & Y \end{tikzcd}$$

to get

$$\begin{aligned}\mathbb{E}(d_X(f(\varepsilon), f(-\varepsilon))^p) &= \mathbb{E}(d_X(\psi^{-1}\circ g(\varepsilon), \psi^{-1}\circ g(-\varepsilon))^p)\\ &\le L_{\psi^{-1}}(1)\mathbb{E}(d_Y(g(\varepsilon), g(-\varepsilon))^p)\\ &\le L_{\psi^{-1}}(1)T_g^p\sum_{i=1}^n \mathbb{E}(d_Y(g(\varepsilon), g(\varepsilon^i)))\\ &\le L_{\psi^{-1}}(1)L_\psi(1)T_g^p\sum_{i=1}^n \mathbb{E}(d_X(g(\varepsilon), g(\varepsilon^i))^p)\end{aligned}$$

as needed. □

The parallelogram inequality for exponent 1 instead of 2 follows from using the triangle inequality on all possible paths for all paths of diagonals. Type $p>1$ is a strengthening of the triangle inequality. For which metric spaces does it hold?

What's an example of a metric space where the inequality doesn't hold with $p>1$? The cube itself (with L^1 distance).

$$n^p \not\lesssim n.$$

I will prove to you that this is the only obstruction: given a metric space that doesn't contain bi-Lipschitz embeddings of arbitrary large cubes, the inequality holds.

We know an alternative inequality involving distance equivalent to type; I can prove it. It is, however, not a satisfactory solution to the Ribe program. There are other situations where we have complete success.

We will prove some things, then switch gears, slow down and discuss Grothendieck's inequality and applications. They will come up in the nonlinear theory later.

2 Bourgain's Theorem implies Ribe's Theorem

2-3-16

We will use the Corson-Klee Lemma 1.1.17.

Proof of Lemma 1.1.17. Suppose $x, y \in X$, $\|x-y\| \ge a$. Break up the line segment from x, y into intervals of length a; let $x = x_0, x_1, \ldots, x_k = y$ be the endpoints of those intervals, with

$$\|x_{i+1} - x_i\| \le a.$$

The **modulus of continuity** is defined as

$$W_f(t) = \sup\left\{\|f(u) - f(v)\| : u, v \in X, \|u-v\| \le t\right\}.$$

Uniform continuity says $\lim_{t\to 0} W_f(t) = 0$. The number of intervals is

$$k \le \frac{\|x-y\|}{a} + 1 \le \frac{2\|x-y\|}{a}.$$

Then

$$\|f(x)-f(y)\| \le \sum_{i=1}^{k} \|f(x_i)-f(x_{i-1})\| \le KW_f(a) \le \frac{2W_f(a)}{a}\|x-y\|,$$

so we can let $L(a)=\frac{2W_f(a)}{a}$. □

2.1 Bourgain's discretization theorem

There are 3 known proofs of Ribe's Theorem.

1. Ribe's original proof, 1987.
2. HK, 1990, a genuinely different proof.
3. Bourgain's proof, a Fourier analytic proof which gives a quantitative version. This is the version we'll prove.

Bourgain uses the Discretization Theorem 1.2.4. There is an amazing open problem in this context.

Saying δ is big says there is a not-too-fine net, which is enough. Therefore we are interested in lower bounds on δ.

Definition 1.2.1 (Discretization modulus)**:** Let X be a finite-dimensional normed space $\dim(X)=n<\infty$. Let the target space Y be an arbitrary Banach space. Consider the unit ball B_X in X. Take a maximal δ-net $\mathcal{N}_\delta$ in B_X. Suppose we can embed $\mathcal{N}_\delta$ into Y via $f:\mathcal{N}_\delta\to Y$. Suppose we know in Y that

$$\|x-y\| \le \|f(x)-f(y)\| \le D\|x-y\|.$$

for all $x,y\in N_\delta$. (We say that $\mathcal{N}_\delta$ embeds with distortion D into Y.)

You can prove using a nice compactness argument that if this holds for δ is small enough, then the entire space X embeds into Y with rough the same distortion. Bourgain's discretization theorem 1.2.4 says that you can choose $\delta=\delta_n$ to be independent of the geometry of X and Y such that if you give a δ-approximation of the unit-ball in the n-dimensional norm, you succeed in embedding the whole space.

I often use this theorem in this way: I use continuous methods to show embedding X into Y requires big distortion; immediately I get an example with a finite object. Let us now make the notion of distortion more precise.

Definition 1.2.2 (Distortion)**:** Suppose $(X,d_X),(Y,d_Y)$ are metric spaces $D\ge 1$. We say that X embeds into Y with **distortion** D if there exists $f:X\to Y$ and $s>0$ such that for all $x,y\in X$,

$$Sd_X(x,y)\le d_Y(f(x),f(y))\le DSd_X(x,y).$$

The infimum over those $D \geq 1$ such that X embeds into Y with distortion is denoted $C_Y(X)$. This is a measure of how far X is being from a subgeometry of Y.

Definition 1.2.3 (**Discretization modulus**)**:** Let be a n-dimensional normed space and Y be any Banach space, $\varepsilon \in (0,1)$. Let $\delta_{X \hookrightarrow Y}(\varepsilon)$ be the supremum over all those $\delta > 0$ such that for every δ-net $\mathcal{N}_\delta$ in B_X,

$$C_Y(\mathcal{N}_\delta) \geq (1-\varepsilon) C_Y(X).$$

Here $B_X := \{x \in X : \|x\| \leq 1\}$.

In other words, the distortion of the δ-net is not much larger than the distortion of the whole space. That is, the discrete δ-ball encodes almost all information about the space when it comes to embedding into Y: If you got $C_Y(\mathcal{N}_\delta)$ to be small, then the distortion of the entire object is not much larger.

Theorem 1.2.4 (Bourgain's discretization theorem)**.** *For every n, $\varepsilon \in (0,1)$, for every X, Y, $\dim X = n$,*

$$\delta_{X \hookrightarrow Y}(\varepsilon) \geq e^{-\left(\frac{n}{\varepsilon}\right)^{Cn}}.$$

Moreover for $\delta = e^{-(2n)^{Cn}}$, we have $C_Y(X) \leq 2C_Y(\mathcal{N}_\delta)$.

Thus there is a δ dependending on the dimension such that in any n-dimensional norm space, the unit ball it encodes all the information of embedding X into *anything else*. It's only a function of the dimension, not of any of the relevant geometry.

The theorem says that if you look at a δ-net in the unit ball, it encodes all the information about X when it comes to embedding into everything else. The amount you have to discretize is just a function of the dimension, and not of any of the other relevant geometry.

Remark 1.2.5: The proof is via a linear operator. All the inequality says is that you can find a *function* with the given distortion. The proof will actually give a *linear operator*.

The best known upper bound is

$$\delta_{X \hookrightarrow Y}\left(\frac{1}{2}\right) \lesssim \frac{1}{n}.$$

The latest progress was 1987, there isn't a better bound yet. You have a month to think about it before you get corrupted by Bourgain's proof.

There is a better bound when the target space is a L^p space.

Theorem 1.2.6 (Gladi, Naor, Shechtman)**.** *For every $p \geq 1$, if $\dim X = n$,*

$$\delta_{X \hookrightarrow L_p}(\varepsilon) \gtrsim \frac{\varepsilon^2}{n^{\frac{5}{2}}}$$

(We still don't know what the right power is.) The case $p = 1$ is important for applications. There are larger classes for spaces where we can write down axioms for where this holds. There are crazy Banach spaces which don't belong to this class, so we're not done. We need more tools to show this: Lipschitz extension theorems, etc.

2.2 Bourgain's Theorem implies Ribe's Theorem

With the "moreover," Bourgain's theorem implies Ribe's Theorem 1.1.9.

Proof of Ribe's Theorem 1.1.9 from Bourgain's Theorem 1.2.4. Let X, Y be Banach spaces that are uniformly homeomorphic. By Corson-Klee 1.1.17, there exists $f : X \to Y$ such that

$$\|x - y\| \geq 1 \implies \|x - y\| \leq \|f(x) - f(y)\| \leq K \|x - y\| .$$

(Apply the Corson-Klee lemma for both f and the inverse.)

In particular, if $R > 1$ and $\mathcal{N}$ is a 1-net in

$$RB_X = \{x \in X : \|x\| \leq R\} ,$$

then $C_Y(\mathcal{N}) \leq K$. Equivalently, for every $\delta > 0$ every δ-net in B_X satisfies $C_Y(\mathcal{N}) \leq K$. If $F \subseteq X$ is a finite dimension subspace and $\delta = e^{-(2 \dim F)^{C \dim F}}$, then by the "moreover" part of Bourgain's Theorem 1.2.4, there exists a linear operator $T : F \to Y$ such that

$$\|x - y\| \leq \|Tx - Ty\| \leq 2K \|x - y\|$$

for all $x, y \in F$. This means that X is finitely representable. □

The motivation for this program comes in passing from continuous to discrete. The theory has many applications, e.g. to computer science whcih cares about finite things. I would like an improvement in Bourgain's Theorem 1.2.4.

First we'll prove a theorem that has nothing to do with Ribe's Theorem. There are lemmas we will be using later. It's an easier theorem. It looks unrelated to metric theory, but the lemmas are relevant.

Chapter 2

Restricted invertibility principle

1 Restricted invertibility principle

1.1 The first restricted invertibility principles

We take basic facts in linear algebra and make things quantitative. This is the lesson of the course: when you make things quantitative, new mathematics appears.

Proposition 2.1.1: If $A : \mathbb{R}^m \to \mathbb{R}^n$ is a linear operator, then there exists a linear subspace $V \subseteq \mathbb{R}^n$ with $\dim(V) = \operatorname{rank}(A)$ such that $A : V \to A(V)$ is invertible.

What's the quantitative question we want to ask about this? Invertibility just says that an inverse exists. Can we find a large subspace where not only is A invertible, but the inverse has small norm?

We insist that the subspace is a coordinate subspace. Let $e_1, \ldots, e_m$ be the standard basis of $\mathbb{R}^m$, $e_j = (0, \ldots, \underbrace{1}_{j}, 0, \ldots)$. The goal is to find a "large" subset $\sigma \subseteq \{1, \ldots, m\}$ such that A is invertible on $\mathbb{R}^\sigma$ where

$$\mathbb{R}^\sigma := \{(x_1, \ldots, x_n) \in \mathbb{R}^m : x_i = 0 \text{ if } i \notin \sigma\}$$

and the norm of $A^{-1} : A(\mathbb{R}^\sigma) \to \mathbb{R}^\sigma$ is small.

A priori this seems a crazy thing to do; take a small multiple of the identity. But we can find conditions that allow us to achieve this goal.

Theorem 2.1.2 (Bourgain-Tzafriri restricted invertibility principle, 1987)**.** thm:btrip *Let $A : \mathbb{R}^m \to \mathbb{R}^m$ be a linear operator such that*

$$\|Ae_j\|_2 = 1$$

for every $j \in \{1, \ldots, m\}$. Then there exist $\sigma \subseteq \{1, \ldots, m\}$ such that

1. *$|\sigma| \geq \frac{cm}{\|A\|^2}$, where $\|A\|$ is the operator norm of A.*

2. A is invertible on $\mathbb{R}^\sigma$ and the norm of $A^{-1} : A(\mathbb{R}^\sigma) \to \mathbb{R}^\sigma$ is at most C' (i.e., $\|AJ_\sigma\|_{S^\infty} \le C'$, to use the notation introduced below).

Here c, C' are universal constants.

Suppose the biggest eigenvalue is at most 100. Then you can always find a coordinate subset of proportional size such that on this subset, A is invertible and the inverse has norm bounded by a universal constant.

All of the proofs use something very amazing.

This proof is from 3 weeks ago. This has been reproved many times. I'll state a theorem that gives better bound than the entire history.

This was extended to rectangular matrices. (The extension is nontrivial.)

Given $V \subseteq \mathbb{R}^m$ a linear subspace with $\dim V = k$ and $A : V \to \mathbb{R}^m$ a linear operator, the singular values of A

$$s_1(A) \ge s_2(A) \ge \cdots \ge s_k(A)$$

are the eigenvalues of $(A^*A)^{\frac{1}{2}}$. We can decompose

$$A = UDV$$

where D is a matrix with $s_i(A)$'s on the diagonal, and U, V are unitary.

Definition 2.1.3: For $p \ge 1$ the **Schatten-von Neumann p-norm** of A is

$$\begin{aligned}\|A\|_{S_p} &:= \left(\sum_{i=1}^k s_i(A)^p\right)^{\frac{1}{p}} \\ &= (\mathrm{Tr}((A^*A)^{\frac{p}{2}}))^{\frac{1}{p}} \\ &= (\mathrm{Tr}((AA^*)^{\frac{p}{2}}))^{\frac{1}{p}}\end{aligned}$$

The cases $p = \infty, 2$ give the operator and Frobenius norm,

$$\begin{aligned}\|A\|_{S_\infty} &= \text{operator norm} \\ \|A\|_{S_2} &= \sqrt{\mathrm{Tr}(A^*A)} = \left(\sum a_{ij}^2\right)^{\frac{1}{2}}.\end{aligned}$$

Exercise 2.1.4: $\|\cdot\|_{S_p}$ is a norm on $\mathcal{M}_{n\times m}(\mathbb{R})$. You have to prove that given A, B,

$$(\mathrm{Tr}([(A+B)^*(A+B)]^{\frac{p}{2}}))^{\frac{1}{p}} \le (\mathrm{Tr}((A^*A)^{\frac{p}{2}}))^{\frac{1}{p}} + (\mathrm{Tr}((B^*B)^{\frac{p}{2}}))^{\frac{1}{p}}.$$

This requires an idea. Note if A, B commute this is trivial. Apparently von Neumann wrote a paper called "Metric Spaces" in the 1930s in which he just proves this inequality and doesn't know what to do with it, so it got forgotten for a while until the 1950s, when Schatten wrote books on applications. When I was a student in grad school, I was taking a class on random matrices. There was two weeks break, I was certain that it was trivial because the professor had not said it was not, and it completely ruined my break though I came up with a different proof of it. It's short, but not trivial: It's not typical linear algebra!. This is like another triangle inequality, which we may need later on.

Spielman and Srivastava have a beautiful theorem.

Definition 2.1.5: Stable rank.
Let $A : \mathbb{R}^m \to \mathbb{R}^n$. The **stable rank** is defined as

$$\operatorname{srank}(A) = \left(\frac{\|A\|_{S_2}}{\|A\|_{S_\infty}} \right)^2 .$$

The numerator is the sum of squares of the singular values, and the denominator is the maximal value. Large stable rank means that many singular values are nonzero, and in fact large on average. Many people wanted to get the size of the subset in the Restricted Invertibility Principle to be close to the stable rank.

Theorem 2.1.6 (Spielman-Srivastava). thm:ss *For every linear operator $A : \mathbb{R}^m \to \mathbb{R}^n$, $\varepsilon \in (0,1)$, there exists $\sigma \subseteq \{1, \ldots, m\}$ with $|\sigma| \ge (1-\varepsilon)\mathit{srank}(A)$ such that*

$$\left\|(AJ_\sigma)^{-1}\right\|_{S_\infty} \lesssim \frac{\sqrt{m}}{\varepsilon \|A\|_{S_2}}.$$

Here, J_σ is the identity function restricted to $\mathbb{R}^\sigma$, $J : \mathbb{R}^\sigma \hookrightarrow \mathbb{R}^m$.

This is stronger than Bourgain-Tzafriri. In Bourgain-Tzafriri the columns were unit vectors.

Proof of Theorem 2.1.2 from Theorem 2.1.6. Let A be as in Theorem 2.1.2. Then $\|A\|_{S_2} = \sqrt{\operatorname{Tr}(A^*A)} = \sqrt{m}$ and $\operatorname{srank}(A) = \frac{m}{\|A\|_{S_\infty}^2}$. We obtain the existence of

$$|\sigma| \ge (1-\varepsilon)\frac{m}{\|A\|_{S_\infty}^2}$$

with $\|(AJ_\sigma)^{-1}\|_{S_\infty} \lesssim \frac{\sqrt{m}}{=}\frac{1}{\varepsilon}$. □

This is a sharp dependence on ε.

The proof introduces algebraic rather than analytic methods; it was eye-opening. Marcus even got sets bigger than the stable rank and looked at $pf\|A\|_{S_2}\|A\|_{S_4}{}^2$, which is much stronger.

1.2 A general restricted invertibility principle

I'll show a theorem that implies all these intermediate theorems. We use (classical) analysis and geometry instead of algebra. What matters is not the ratio of the norms, but the tail of the distribution of $s_1(A)^2, \ldots, s_m(A)^2$.

Theorem 2.1.7. thm:gen-srank *Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator. If $k < \operatorname{rank}(A)$ then there exist $\sigma \subseteq \{1, \ldots, m\}$ with $|\sigma| = k$ such that*

$$\left\|(AJ_\sigma)^{-1}\right\|_{S_\infty} \lesssim \min_{k < r \le \operatorname{rank}(A)} \sqrt{\frac{mr}{(r-k)\sum_{i=r}^m s_i(A)^2}}.$$

You have to optimize over r. You can get the ratio of L_p norms from the tail bounds. This implies all the known theorems in restricted invertibility. The subset can be as big as you want up to the rank, and we have sharp control in the entire range. This theorem generalizes Spielman-Srivasta (Theorem 2.1.6), which had generalized Bourgain-Tzafriri (Theorem 2.1.2). 2-8-16

Now we will go backwards a bit, and talk about a less general result. After Theorem 2.1.6, a subsequent theorem gave the same theorem but instead of the stable rank, used something better.

Theorem 2.1.8 (Marcus, Spielman, Srivastava). thm:mss4 *If*

$$k < \frac{1}{4}\left(\frac{\|A\|_{S_2}}{\|A\|_{S_4}}\right)^4,$$

there exists $\sigma \subseteq \{1, \dots, m\}$, $|\sigma| = k$ *such that*

$$\left\|(AJ_\sigma)^{-1}\right\|_{S_\infty} \lesssim \frac{\sqrt{m}}{\|A\|_{S_2}}.$$

A lot of these quotients of norms started popping up in people's results. The correct generalization is the following notion.

Definition 2.1.9: For $p > 2$, define the **stable pth rank** by

$$\operatorname{srank}_p(A) = \left(\frac{\|A\|_{S_2}}{\|A\|_{S_p}}\right)^{\frac{2p}{p-2}}.$$

Exercise 2.1.10: Show that if $p \geq q > 2$, then

$$\operatorname{srank}_p(A) \leq \operatorname{srank}_q(A).$$

(Hint: Use Hölder's inequality.)

Now we would like to prove how Theorem 2.1.7 generalizes the previously listed results:

Proof of Generalizability of Theorem 2.1.7. Using Hölder's inequality with $\frac{p}{2}$,

$$\begin{aligned}
\|A\|_{S_2}^2 &= \sum_{j=1}^{m} s_j(A)^2 \\
&= \sum_{j=1}^{r-1} s_j(A)^2 + \sum_{j=r}^{m} s_j(A)^2 \\
&\leq (r-1)^{1-\frac{2}{p}} \left(\sum_{j=1}^{r-1} s_j(A)^p\right)^{\frac{2}{p}} + \sum_{j=r}^{m} s_j(A)^2
\end{aligned}$$

$$\begin{aligned}
&\le (r-1)^{1-\frac{2}{p}} \|A\|_{S_p}^2 + \sum_{j=r}^m s_j(A)^2 \\
\sum_{j=r}^m s_j(A)^2 &\ge \|A\|_{S_2}^2 \left(1-(r-1)^{-\frac{2}{p}} \frac{\|A\|_{S_p}^2}{\|A\|_{S_2}^2}\right) \\
&= \|A\|_{S_2}^2 \left(1-\left(\frac{r-1}{\mathrm{srank}_p(A)}\right)^{1-\frac{2}{p}}\right)
\end{aligned}$$

Now we can plug the previous calculation into Theorem 2.1.7 to demonstrate the way the new theorem generalizes the previous results:

$$\begin{aligned}
\left\|(AJ_\sigma)^{-1}\right\| &\lesssim \min_{k+1\le r\le \mathrm{rank}(A)} \sqrt{\frac{mr}{(r-k)\|A\|_{S_2}^2\left(1-\left(\frac{r-1}{\mathrm{srank}_p(A)}\right)^{1-\frac{2}{p}}\right)}} \\
&= \frac{\sqrt{m}}{\|A\|_\infty} \min_{k+1\le r\le \mathrm{rank}(A)} \sqrt{\frac{r}{(r-k)\left(1-\left(\frac{r-1}{\mathrm{srank}_p(A)}\right)^{1-\frac{2}{p}}\right)}}
\end{aligned}$$

This equation implies all the earlier theorems. □

To optimize, fix the stable rank, differentiate in r, and set to 0. All theorems in the literature follow from this theorem; in particular, we get all the bounds we got before. There was nothing special about the number 4 in Theorem 2.1.8; this is about the distribution of the eigenvalues.

2 Ky Fan maximum principle

sec:kf We'll be doing linear algebra. It's mostly mechanical, except we'll need this lemma.

Lemma 2.2.1 (Ky Fan maximum principle)**.** *lem:kf Suppose that $P:\mathbb{R}^m \to \mathbb{R}^m$ is a rank k orthogonal projection. Then*

$$\mathrm{Tr}(A^*AP) \le \sum_{i=1}^k s_i(A)^2$$

where $s_i(A)$ are the singular values, i.e., $s_i(A)^2$ are the eigenvalues of $B := A^*A$.

This material was given in class on 2-15. This lemma follows from the following general theorem.

Theorem 2.2.2 (Convex function of dot products acheives maximum at eigenvectors)**.** *thm:eignmax Let $B:\mathbb{R}^n\to\mathbb{R}^n$ be symmetric, and let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function. Then for every orthonormal basis $u_1,\dots,u_n\in\mathbb{R}^n$, there exists a permutation π such that*

$$\text{eq:kf0}\, f(\langle Bu_1,u_1\rangle,\dots,\langle Bu_n,u_n\rangle) \le f(\lambda_{\pi(1)},\dots,\lambda_{\pi(n)}) \tag{2.1}$$

Essentially, we're saying using the eigenbasis maximizes the convex function.

Remark 2.2.3: We can weaken the condition on f to the following: for every $i < j$, $t \mapsto f(x_1, \ldots, x_i + t, x_{i+1}, \ldots, x_{j-1}, x_j - t, x_{j+1}, \ldots, x_n)$ is convex as a function of t. If f is smooth, this is equivalent to the second derivative in t being ≥ 0:

$$\text{eq:kf1}\frac{\partial^2 f}{\partial x_i^2} + \frac{\partial^2 f}{\partial x_j^2} - 2\frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0 \tag{2.2}$$

In other words, you just need that the Hessian is positive semidefinite. (Above, we wrote the determinant of the Hessian on the $x_i - x_j$ plane. This being true for all pairs i, j is the same as the Hessian being positive definite.)

Proof. We may assume without loss of generality that

1. f is smooth. If not, convolute with a good kernel.
2. Strict inequality holds:
$$\frac{\partial^2 f}{\partial x_i^2} + \frac{\partial^2 f}{\partial x_j^2} - 2\frac{\partial^2 f}{\partial x_i \partial x_j} > 0.$$
To see this, note we can take any $\varepsilon > 0$ and perturb $f(x)$ into $f(x) + \varepsilon \|x\|_2^2$. The inequality to prove (2.1) is also perturbed by this slight change, and taking $\varepsilon \to 0$ gives the desired inequality.

Now let $u_1, \ldots, u_n$ be an orthonormal basis at which $f(\langle Bu_1, u_1\rangle, \ldots, \langle Bu_n, u_n\rangle)$ attains its maximum. Then for u_i, u_j, we want to rotate in the $i - j$ plane by angle θ. Since u_i, u_j span a two dimensional subspace, recall the 2-dimensional rotation matrix. Let

$$R_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}; u_{i;j} = \begin{bmatrix} u_i \\ u_j \end{bmatrix}$$

Multiplying, we get

$$R_\theta u_{i;j} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \begin{bmatrix} u_i \\ u_j \end{bmatrix} = \begin{bmatrix} \cos(\theta)u_i + \sin(\theta)u_j \\ \sin(\theta)u_i - \cos(\theta)u_j \end{bmatrix} = \begin{bmatrix} (R_\theta u_{i;j})_1 \\ (R_\theta u_{i;j})_2 \end{bmatrix}$$

Then, we replace f with $g(\theta) =$

$$f\left(\langle Bu_1, u_1\rangle, \ldots, \langle B\left(R_\theta u_{i;j}\right)_1, (R_\theta u_{i;j})_1\rangle, \langle B\left(R_\theta u_{i;j}\right)_2, (R_\theta u_{i;j})_2\rangle, \ldots, \langle Bu_n, u_n\rangle\right)$$

where we keep all other dot products the same. By assumption, g attains its maximum at $\theta = 0$, so $g'(0) = 0, g''(0) \leq 0$. Expanding out the rotated dot products explicitly in $g(\theta)$, we get that the ith argument is

$$\cos^2(\theta)\langle Bu_i, u_i\rangle + \sin^2(\theta)\langle Bu_j, u_j\rangle + \sin(2\theta)\langle Bu_i, u_j\rangle$$

and the jth argument is

$$\sin^2(\theta)\langle Bu_i, u_i\rangle + \cos^2(\theta)\langle Bu_j, u_j\rangle - \sin(2\theta)\langle Bu_i, u_j\rangle$$

Then we can mechanically take the derivatives at $\theta = 0$ to get

$$\begin{aligned} 0 &= g'(0) = 2\langle Bu_i, u_j\rangle(f_{x_i} - f_{x_j}) \\ 0 &\geq g''(0) = 2\left(\langle Bu_j, u_j\rangle - \langle Bu_i, u_i\rangle\right) \underbrace{(f_{x_i} - f_{x_j})}_{=0 \text{ if } \langle Bu_i, u_j\rangle \neq 0} + 4\langle Bu_i, u_j\rangle^2 \underbrace{\left(f_{x_i x_i} + f_{x_j x_j} - 2f_{x_i x_j}\right)}_{\geq 0}. \end{aligned}$$

This implies that for all $i \neq j$ $\langle Bu_i, u_j\rangle = 0$, which implies that for all i, $Bu_i = \mu_i u_i$ for some μ_i. Thus any function applied to a vector of dot products is maximized at eigenvalues. □

Exercise 2.2.4: exr:kf If $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the conditions in Theorem 2.2.2 and $(u_1, \ldots, u_n)$, $(v_1, \ldots, v_n)$ are two orthonormal bases of $\mathbb{R}^n$, then for every $A : \mathbb{R}^n \to \mathbb{R}^n$, there exists $\pi \in S_n$, $(\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n$ such that

$$f(\langle Au_1, v_1\rangle, \langle Au_2, v_2\rangle, \ldots, \langle Au_n, v_n\rangle) \leq f(\varepsilon_1 s_{\pi(1)}(A), \ldots, \varepsilon_n s_{\pi(n)}(A))$$

Show that choosing u, v as the singular vectors maximizes f (over all pairs of orthonormal bases).

To solve this problem, you can rotate both vectors in the same direction and take derivatives, and also rotate them in opposite directions and take derivatives to get enough information to prove that the singular values are the maximum.

Essentially, a lot of the inequalities you find in books follow from this. For instance, if you want to prove that the Schatten p-norm is a norm, it follows directly from this fact.

Corollary 2.2.5. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ that is invariant under premutations and sign:*

$$\|(x_1, \ldots, x_n)\| = \|(\varepsilon_1 x_{\pi(1)}, \ldots, \varepsilon_n x_{\pi(n)})\|$$

for all $\varepsilon \in \{\pm 1\}^n$ and $\pi \in S_n$ (In the literature, we call this a ***symmetric norm****). This induces a norm on matrices $M_{m\times n}(\mathbb{R})$ with*

$$\|A\| = \|(s_{\pi(1)}(A), \ldots, s_{\pi(n)}(A)\|)$$

Then the triangle inequality holds for matrices A, B:

$$\|A + B\| \leq \|A\| + \|B\|.$$

Proof. We have by Exercise 2.2.4

$$\begin{aligned} \|A + B\| &= \max_{(u_i)\perp, (v_i)\perp} \|(\langle (A+B)u_i, v_i\rangle)_{i=1}^n\| \\ &\leq \max_{(u_i)\perp, (v_i)\perp} \|(\langle (Au_i, v_i\rangle)_{i=1}^n\| + \max_{(u_i)\perp, (v_i)\perp} \|(\langle Bu_i, v_i\rangle)_{i=1}^n\| \\ &\leq \|A\| + \|B\|. \end{aligned}$$

□

Remember Theorem 2.2.2! For many results, you simply need to apply the right convex function to get the result.

Our lemma follows from setting $f(x) = \sum_{i=1}^k x_i$.

Proof of Ky Fan Maximum Principle (Lemma 2.2.1). Take an orthonormal basis $u_1, \ldots, u_n$ of P such that $u_1, \ldots, u_k$ is a basis of the range of P. Then

$$\mathrm{Tr}(BP) = \sum_{j=1}^k \langle Be_j, e_j \rangle \le \sum_{i=1}^k s_i(B) = \sum_{i=1}^k s_i(A)^2$$

□

3 Finding big subsets

We'll present 4 lemmas for finding big subsets with certain properties. We'll put them together at the end.

3.1 Little Grothendieck inequality

Theorem 2.3.1 (Little Grothendieck inequality)**.** thm:lgi *Fix $k, m, n \in \mathbb{N}$. Suppose that $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear operator. Then for every $x_1, \ldots, x_k \in \mathbb{R}^m$,*

$$\sum_{r=1}^k \|Tx_r\|_2^2 \le \frac{\pi}{2} \|T\|_{\ell_\infty^m \to \ell_2^n}^2 \sum_{r=1}^k x_{ri}^2$$

for some $i \in \{1, \ldots, m\}$ where $x_r = (x_{r1}, \ldots, x_{rm})$.

Later we will show $\frac{\pi}{2}$ is sharp.

If we had only 1 vector, what does this say?

$$\|Tx_1\|_2 \le \sqrt{\frac{\pi}{2}} \|T\|_{\ell_\infty^m \to \ell_2^n} \|x_1\|_\infty$$

We know the inequality is true for $k = 1$ with constant 1, by definition of the operator norm. The theorem is true for arbitrary many vectors, losing an universal constant ($\frac{\pi}{2}$). After we see the proof, the example where $\frac{\pi}{2}$ is attained will be natural.

We give Grothendieck's original proof.

The key claim is the following.

Claim 2.3.2. clm:lgi

$$\text{eq:lgi1} \sum_{j=1}^m \left(\sum_{r=1}^k (T^*Tx_r)_j^2 \right)^{\frac{1}{2}} \le \sqrt{\frac{\pi}{2}} \|T\|_{\ell_\infty^m \to \ell_2^n} \left(\sum_{r=1}^k \|Tx_r\|^2 \right)^{\frac{1}{2}}. \tag{2.3}$$

Proof of Theorem 2.3.1. Assuming Claim 2.3.2, we prove the theorem.

$$\begin{aligned}
\sum_{r=1}^k \|Tx_r\|_2^2 &= \sum_{r=1}^k \langle Tx_r, Tx_r\rangle \\
&= \sum_{r=1}^k \langle x_r, T^*Tx_r\rangle \\
&= \sum_{r=1}^k \sum_{j=1}^m x_{rj}(T^*Tx_r)_j \\
&\le \sum_{j=1}^m \left(\sum_{r=1}^k x_{rj}^2\right)^{\frac12} \left(\sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12} && \text{by Cauchy-Schwarz} \\
&\le \left(\max_{1\le j\le m} \left(\sum_{r=1}^k x_{rj}^2\right)^{\frac12}\right) \left(\sum_{j=1}^m \sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12} \\
&\le \max_{1\le j\le m} \left(\sum_{i=1}^k x_{ij}^2\right)^{\frac12} \sqrt{\frac{\pi}{2}}\, \|T\|_{\ell_\infty^m\to\ell_2^n} \left(\sum_{r=1}^k \|Tx_r\|_2^2\right)^{\frac12} \\
\sum_{r=1}^k \|Tx_r\|_2^2 &\le \frac{\pi}{2}\|T\|_{\ell_\infty^m\to\ell_2^n}^2 \max_j \sum_{r=1}^k x_{ij}^2.
\end{aligned}$$

We bounded by a square root of the multiple of the same term, a bootstrapping argument. In the last step, divide and square. □

Proof of Claim 2.3.2. Let $g_1, \ldots, g_k$ be iid standard Gaussian random variables. For every fixed $j \in \{1, \ldots, m\}$,

$$\sum_{r=1}^k g_r (T^*Tx_r)_j.$$

This is a Gaussian random variable with mean 0 and variance $\sum_{r=1}^k (T^*Tx_r)_j^2$. Taking the expectation,[1]

$$\mathbb{E}\left|\sum_{r=1}^k g_r(T^*Tx_r)_j\right| = \left(\sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12}\sqrt{\frac{2}{\pi}}.$$

Sum these over j:

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^m \left|T^*(\sum_{r=1}^k g_r Tx_r)_j\right|\right] &= \sqrt{\frac{2}{\pi}} \sum_{j=1}^m \left(\sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12} \\
\sum_{j=1}^m \left(\sum_{r=1}^k (T^*Tx_r)_j^2\right)^{\frac12} &= \sqrt{\frac{\pi}{2}}\mathbb{E}\left[\sum_{j=1}^m \left|T^*\sum_{r=1}^k g_r(Tx_r)_j\right|\right]. \text{eq:lgi2} \qquad (2.4)
\end{aligned}$$

[1] $\sqrt{\frac{1}{2\pi}}\int_{-\infty}^{\infty}|x|e^{-\frac{x^2}{2}} = -2\sqrt{\frac{1}{2\pi}}[e^{-\frac{x^2}{2}}]_0^\infty = \sqrt{\frac{2}{\pi}}$

Define a random sign vector $z \in \{\pm 1\}^m$ by

$$z_j = \text{sign}\left(\left(T^* \sum_{r=1}^k g_r T x_r\right)_j\right)$$

Then

$$\begin{aligned}
\sum_{j=1}^m \left|(T^* \sum_{r=1}^k gTx_r)_j\right| &= \left\langle z, T^* \sum_{r=1}^k g_r T x_r \right\rangle \\
&= \left\langle Tz, \sum_{r=1}^k g_r T x_r \right\rangle \\
&\le \|Tz\|_2 \left\| \sum_{r=1}^k g_r T x_r \right\|_2 \\
&\le \|T\|_{\ell_\infty^m \to \ell_2^n} \left\| \sum_{r=1}^k g_r T x_r \right\|_2
\end{aligned}$$

This is a pointwise inequality. Taking expectations and using Cauchy-Schwarz,

eq:lgi3
$$\mathbb{E}\left[\sum_{j=1}^m \left|\left(T^* \sum_{r=1}^k g_r T x_r\right)_j\right|\right] \le \|T\|_{\ell_\infty^m \to \ell_2^n} \left(\mathbb{E}\left\|\sum_{r=1}^k g_r T x_r\right\|_2^2\right)^{\frac{1}{2}}. \tag{2.5}$$

What is the second moment? Expand:

eq:lgi4
$$\mathbb{E}\left\|\sum_{r=1}^k g_i T x_r\right\|_2^2 = \mathbb{E}\left[\sum_{ij} g_i g_j \langle Tx_i, Tx_j \rangle\right] = \sum_{r=1}^k \|Tx_r\|_2^2. \tag{2.6}$$

Chaining together (2.4), (2.5), (2.6) gives the result. □

Why use the Gaussians? The identity characterizes the Gaussians using rotation invariance. Using other random variables gives other constants that are not sharp.

There will be lots of geometric lemmas:

- A fact about restricting matrices.
- Another geometric argument to give a different method for selecting subsets.
- A combinatorial lemma for selecting subsets.

Finally we'll put them together in a crazy induction.

2-10-16: We were in the process of proving three or four subset selection principles, which we will somehow use to prove the RIP.

The little Grothendieck inequality (Theorem 2.3.1) is part of an amazing area of mathematics with many applications. It's little, but very useful. The proof is really Grothendieck's original proof, re-organized. For completeness, we'll show the fact that the inequality is sharp (cannot be improved).

3.1.1 Tightness of Grothendieck's inequality

Corollary 2.3.3. $\sqrt{\pi/2}$ *is the best constant in Theorem 2.3.1.*

From the proof, we reverse engineer vectors that make the inequality sharp. They are given in the following example.

Example 2.3.4: Let $g_1, g_2, \ldots, g_k$ be iid Gaussians on the probability space (Ω, P). Let $T : L_\infty(\Omega, P) \to \ell_2^k$ be

$$Tf = (\mathbb{E}[fg_1], \ldots, \mathbb{E}[fg_k]).$$

Let $x_r \in L_\infty(\Omega, P)$,

$$x_r = \frac{g_r}{\left(\sum_{i=1}^k g_i^2\right)^{\frac{1}{2}}}.$$

Proof. Le $g_1, \ldots, g_k; T; x_1, \ldots x_k$ be as in the example. Note the x_r are nothing more than vectors on the k-dimensional unit sphere, so they are bounded functions on the measure space Ω. We can also write

$$\text{eq:sum-xr} \sum_{r=1}^k x_r(\omega)^2 = \sum_{r=1}^k \frac{g_r(\omega)^2}{\sum_{i=1}^r g_i(\omega)^2} = 1 \tag{2.7}$$

We use the Central Limit Theorem in order to replace the Ω by a discrete space. Let $\varepsilon_{r,i}$ be ± 1 random variables. Then letting

$$g_r = \frac{\varepsilon_{r,1} + \ldots + \varepsilon_{r,N}}{\sqrt{N}}$$

instead, we have that g_r approaches a standard Gaussian in distribution, so the statements we make will be asymptotically true. With this discretization, the random variables $\{g_r\}$ live in $\Omega = \{\pm 1\}^{NK}$. So $L_\infty(\Omega) = l_\infty^{2^{NK}}$, which is in a large but finite dimension. So ω will really be a coordinate in Ω.

Now we show two things; they are nothing more than computations.

1. $\|T\|_{L_\infty(\Omega,\mathbb{P}) \to l_2^k} = \sqrt{2/\pi}$,

2. We also show $\sum_{r=1}^k \|Tx_r\|_2^2 \xrightarrow{k\to\infty} 1$.

From (2.7) and the 2 items, the little Grothendieck inequality is sharp in the limit.

For (1), we have

$$\begin{aligned}
\|T\|_{\ell^\infty\to\ell^2} &= \sup_{\|f\|_\infty\le 1}\left(\sum_{r=1}^k \mathbb{E}\left[fg_r\right]^2\right)^{1/2} \\
&= \sup_{\|f\|_\infty\le 1}\sup_{\sum_{r=1}^k\alpha_r^2=1}\sum_{r=1}\alpha_r\mathbb{E}\left[fg_r\right] \\
&= \sup_{\sum_{r=1}^k\alpha_r^2=1}\sup_{\|f\|_\infty\le 1}\mathbb{E}\left[f\sum_{i=1}^k\alpha_r g_r\right] \\
&= \sup_{\sum_{r=1}^k}\mathbb{E}\left|\sum_{r=1}^k\alpha_r g_r\right| = \mathbb{E}|g_1| = \sqrt{\frac{2}{\pi}}
\end{aligned} \tag{2.8}$$

as we claimed, since $\|\alpha\|_2 = 1$ implies $\sum_{r=1}^k\alpha_r g_r$ is also a gaussian.

Now for (2),

$$\begin{aligned}
\sum_{r=1}^k\|Tx_r\|_2^2 &= \sum_{r=1}^k\left(\mathbb{E}\left[\frac{g_r^2}{\left(\sum_{i=1}^k g_i^2\right)^{1/2}}\right]\right)^2 \\
&= K\left(\mathbb{E}\left[\frac{g_1^2}{\left(\sum_{i=1}^k g_i^2\right)^{1/2}}\right]\right)^2 \\
&= K\left(\frac{1}{K}\mathbb{E}\left[\sum_{r=1}^k\frac{g_r^2}{\left(\sum_{i=1}^k g_i^2\right)^{1/2}}\right]\right)^2 \\
&= \frac{1}{K}\left(\mathbb{E}\left[\left(\sum_{i=1}^k g_i^2\right)^{1/2}\right]\right)^2
\end{aligned} \tag{2.9}$$

and you can use Stirling to finish. This is just a χ^2-distribution.

In this case $\mathbb{E}\frac{g_1g_2}{\left(\sum_i g_i^2\right)^{1/2}} = \mathbb{E}\frac{g_1(-g_2)}{\left(\sum_i g_i^2\right)^{1/2}}$. Also note that if $(g_1,\dots,g_k)\in\mathbb{R}^k$ is a standard Gaussian, then $\frac{(g_1,\dots,g_k)}{\left(\sum_{i=1}^k g_i^2\right)^{1/2}}$ and $\left(\sum_{i=1}^k g_i^2\right)^{1/2})$ are independent. In other words, the length and angle are independent: This is just polar coordinates, you can check this. □

Now, how does this relate to the Restricted Invertibility Problem?

3.2 Pietsch Domination Theorem

Theorem 2.3.5 (Pietsch Domination Theorem). thm:pdt *Fix $m,n\in\mathbb{N}$ and $M>0$. Suppose that $T:\mathbb{R}^m\to\mathbb{R}^n$ is a linear operator such that for every $x_1,\dots,x_k\in\mathbb{R}^m$ have*

$$\text{eq:pdt1}\quad\left(\sum_{r=1}^k\|Tx_r\|_2^2\right)^{1/2}\le M\max_{1\le j\le m}\left(\sum_{r=1}^k x_{rj}^2\right)^{1/2} \tag{2.10}$$

Then there exist $\mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m$ with $\mu_1 \geq 0$ and $\sum_{i=1}^m \mu_i = 1$ such that for every $x \in \mathbb{R}^m$

$$\text{eq:pdt2} \|Tx\|_2 \leq M \left(\sum_{i=1}^M \mu_i \|x_i\|^2 \right)^{1/2} \tag{2.11}$$

The theorem says that you can come up with a probability measure such that the norm of T as an operator as a standard norm from l_∞ to l_2 (?), is bounded by M.

Remark 2.3.6: The theorem really an iff: (2.11) is a stronger statement than (2.10), and in fact they are equivalent.

Proof. Define $K \subseteq \mathbb{R}^m$ with

$$K = \left\{ y \in \mathbb{R}^m : y_i = \sum_{r=1}^k \|Tx_r\|_2^2 - M^2 \sum_{r=1}^m x_{ri}^2 \text{ for some } k, x_1, \dots, x_k \in \mathbb{R}^m \right\}$$

Basically we cleverly select a convex set. Every n-tuple of vectors in $\mathbb{R}^m$ gives you a new vector in $\mathbb{R}^m$. Let's check that K is convex. We have to check if two vectors $y, z \in K$, then all points on the line between them are in K. $y, z \in K$ means that there exist $(x_i)_{i=1}^k$, $(w_i)_{i=1}^l$,

$$y_i = \sum_{r=1}^k \|Tx_r\|_2^2 - M^2 \sum_{r=1}^m x_{ri}^2$$
$$z_i = \sum_{r=1}^l \|Tw_r\|_2^2 - M^2 \sum_{r=1}^l w_{ri}^2$$

for all i. Then $\alpha y_i + (1-\alpha) z_i$ comes from $(\sqrt{\alpha}x_1, \dots, \sqrt{\alpha}x_k, \sqrt{1-\alpha}w_1, \dots \sqrt{1-\alpha}w_k)$. So by design, K is a convex set.

Now, the assumption of the theorem says that

$$\left(\sum_{r=1}^k \|Tx_r\|_2^2 \right)^{1/2} \leq M \max_{1 \leq j \leq m} \left(\sum_{r=1}^k x_{rj}^2 \right)^{1/2}$$

which implies

$$\|Tx_r\|_2^2 - M^2 \max_{1 \leq j \leq m} \sum_{r=1}^m x_{rj}^2 \leq 0$$

which implies $K \cap (0, \infty)^m = \emptyset$. By the hyperplane separation theorem (for two disjoint convex sets in $\mathbb{R}^m$ with at least one compact, there is a hyperplane between them), there exists $0 \neq \mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m$ with

$$\langle \mu, y \rangle \leq \langle \mu, z \rangle$$

for all $y \in K$ and $z \in (0, \infty)^m$. By renormalizing, we may assume $\sum_{i=1}^m \mu_i = 1$. Moreover μ cannot have any strictly negative coordinate: Otherwise you could take z to have arbitrarily large value at a strictly negative coordinate with zeros everywhere else, implying $\langle u, z\rangle$ is no longer bounded from below, a contradiction. Therefore, μ is a probability vector and $\langle \mu, z\rangle$ can be arbitrarily small. So for every $y \in K$, $\sum_{i=1}^m \mu_i y_i \le 0$. Write

$$y_i = \|Tx\|_2^2 - M^2 \|x_i\|^2 \in K.$$

Expanding this out,

$$\|Tx\|_2^2 - M^2 \sum_{i=1}^n \mu_i \|x_i\|^2 \le 0,$$

which is exactly what we wanted. □

3.3 A projection bound

Lemma 2.3.7. lem:projbound *$m, n \in \mathbb{N}$, $\varepsilon \in (0,1)$, $T : \mathbb{R}^n \to \mathbb{R}^m$ a linear operator. Then $\exists \sigma \subset \{1, \ldots, m\}$ with $|\sigma| \ge (1-\varepsilon)m$ such that*

$$\|Proj_{\mathbb{R}^\sigma} T\|_{S_\infty} \le \sqrt{\frac{\pi}{2\varepsilon m}} \|T\|_{l_2^n \to l_1^m}$$

We will find ways to restrict a matrix to a big submatrix. We won't be able to control its operator norm, but we will be able to control the norm from l_2^n to l_1^m. Then we pass to a further subset, which this becomes an operator norm on, which is an improvement which Grothendieck gave us. This is the first very useful tool to start finding big submatrices.

Proof. We have $T : l_2^n \to l_1^m$, $T^* : l_\infty^m \to l_2^n$. Now some abstract nonsense gives us that for Banach spaces, the norm of an operator and its adjoint are equal, i.e. $\|T\|_{l_2^n \to l_1^m} = \|T^*\|_{l_\infty^m \to l_2^n}$. This statement follows from the Hahn-Banach theorem (come see me if you haven't seen this before, I'll tell you what book to read). From the Little Grothendieck inequality (Theorem 2.3.1), T^* satisfies the assumption of the Pietsch domination theorem 2.3.5 with $M = \sqrt{\frac{\pi}{2}} \|T\|_{l_2^n \to l_1^m}$ (we're applying it to T^*). By the theorem, there exists a probability vector $(\mu_1, \ldots, \mu_m)$ such that for every $y \in \mathbb{R}^m$,

$$\|T^* y\|_2 = M \left(\sum_{i=1}^m \mu_i y_i^2 \right)^{1/2}$$

with $M = \sqrt{\frac{\pi}{2}} \|T\|_{l_2^n \to l_1^m}$. Define $\sigma = \left\{ i \in \{1, \ldots, m\} : \mu_i \le \frac{1}{m\varepsilon} \right\}$; then $|\sigma| \ge (1-\varepsilon)m$ by Markov's inequality. We can also see this by writing

$$1 = \sum_{i=1}^m \mu_i = \sum_{i \in \sigma} \mu_i + \sum_{i \notin \sigma} \mu_i > \sum_{i \in \sigma} \mu_i + \frac{m - |\sigma|}{m\varepsilon}$$

which follows since for $j \notin \sigma$, $\mu_j > \frac{1}{m\varepsilon}$. Continuing,

$$\frac{m\varepsilon - m + |\sigma|}{m\varepsilon} \ge \sum_{i \in \sigma} \mu_i$$
$$|\sigma| \ge (m\varepsilon) \sum_{i \in \sigma} \mu_i + m(1-\varepsilon)$$

Then, because μ is a probability distribution, $(m\varepsilon)\sum_{i\in\sigma}\mu_i \geq 0$ and we have

$$|\sigma| \geq m(1-\varepsilon)$$

Now take $x \in \mathbb{R}^n$ and choose $y \in \mathbb{R}^m$ with $\|y\|_2 = 1$. Then

$$\begin{aligned}\langle y, \mathrm{Proj}_{\mathbb{R}^\sigma} Tx\rangle^2 &= \langle T^*\mathrm{Proj}_{\mathbb{R}^\sigma} y, x\rangle^2 \leq \|T^*\mathrm{Proj}_{\mathbb{R}^\sigma} y\|_2^2 \cdot \|x\|_2^2 \\ &\leq \frac{\pi}{2}\|T\|_{l_2^n\to l_1^m}\left(\sum_{i\in\sigma}\mu_i y_i^2\right)\|x\|_2^2 \leq \frac{\pi}{2}\|T\|^2_{l_2^n\to l_1^m}\frac{1}{m\varepsilon}\|x\|_2^2\end{aligned}$$

by Cauchy-Schwarz. Taking square roots gives the desired result. □

In the previous proof, we used a lot of duality to get an interesting subset.

Remark 2.3.8: In Lemma 2.3.7, I think that either the constant $\pi/2$ is sharp (no subset are bigger; it could come from the Gaussians), or there is a different constant here. If the constant is 1, I think you can optimize the previous argument and get the constant to be arbitrarily close to 1, which would have some nice applications: In other words, getting $\sqrt{\frac{\pi}{2\varepsilon m}}$ as close to 1 as possible would be good. I didn't check before class, but you might want to check if you can carry out this argument using the Gaussian argument we made for the sharpness of $\frac{\pi}{2}$ in Grothendieck's inequality (Theorem 2.3.1). It's also possible that there is a different universal constant.

3.4 Sauer-Shelah Lemma

Now we will give another lemma which is very easy and which we will use a lot.

Lemma 2.3.9 (Sauer-Shelah). lem:saushel *Take integers $m, n \in \mathbb{N}$ and suppose that we have a large set $\Omega \subseteq \{\pm 1\}^n$ with*

$$|\Omega| > \sum_{k=0}^{m-1}\binom{n}{k}$$

Then $\exists \sigma \subseteq \{1, \ldots, n\}$ such that with $|\sigma| = m$, if you project onto $\mathbb{R}^\sigma$ the set of vectors, you get the entire cube: $Proj_{\mathbb{R}^\sigma}(\Omega) = \{\pm 1\}^\sigma$.[2] *For every $\varepsilon \in \{\pm 1\}^\sigma$, there exists $\delta = (\delta_1, \ldots, \delta_n) \in \Omega$ such that $\delta_j = \varepsilon_j$ for $j \in \sigma$.*

Note that Lemma 2.3.9 is used in the proof of the Fundamental Theorem of Statistical Learning Theory.

Proof. We want to prove by induction on n. First denote the **shattering set**

$$\mathrm{sh}(\Omega) = \{\sigma \subseteq \{1, \ldots, n\} : \mathrm{Proj}_{\mathbb{R}^\sigma}\Omega = \{\pm 1\}^\sigma\}$$

[2] I.e., the **VC dimension** of Ω is $\geq m$.

The claim is that the number of sets shattered by a given set is $|\text{sh}(\Omega)| \geq |\Omega|$. The empty set case is trivial. What happens when $n = 1$? $\Omega \subset \{-1, 1\}$, and thus the set is shattered. Assume that our claim holds for n, and now set $\Omega \subseteq \{\pm 1\}^{n+1} = \{\pm 1\}^n \times \{\pm 1\}$. Define

$$\Omega_+ = \{\omega \in \{\pm 1\}^n : (\omega, 1) \in \Omega\}$$

$$\Omega_- = \{\omega \in \{\pm 1\}^n : (\omega, -1) \in \Omega\}$$

Then, letting $\tilde{\Omega}_+ = \{(\omega, 1) \in \{\pm 1\}^{n+1} : \omega \in \Omega_+\}$ and $\tilde{\Omega}_-$ similarly, we have $|\Omega| = |\tilde{\Omega}_+| + |\tilde{\Omega}_-| = |\Omega_+| + |\Omega_-|$. By our inductive step, we have $\text{sh}(\Omega_+) \geq |\Omega_+|$ and $\text{sh}(\Omega_-) \geq |\Omega_-|$. Note that any subset that shatters Ω_+ also shatters Ω, and likewise for Ω_-. Note that if a set Ω' shatters both of them, we are allowed to add on an extra coordinate to get $\Omega' \times \{\pm 1\}$ which shatters Ω. Therefore,

$$\text{sh}(\Omega_+) \cup \text{sh}(\Omega_-) \cup \{\sigma \cup \{n+1\} : \sigma \in \text{sh}(\Omega_+) \cap \text{sh}(\Omega_-)\} \subseteq \text{sh}(\Omega)$$

where the last union is disjoint since the dimensions are different. Therefore, we can now use this set inclusion to complete the induction using the principle of inclusion-exclusion:

$$\begin{aligned}
|\text{sh}(\Omega)| &\geq |\text{sh}(\Omega_+) \cup \text{sh}(\Omega_-)| + |\text{sh}(\Omega_+) \cap \text{sh}(\Omega_-)| \qquad\qquad \text{(disjoint sets)} \\
&= |\text{sh}(\Omega_+)| + |\text{sh}(\Omega_-)| - |\text{sh}(\Omega_+) \cap \text{sh}(\Omega_-)| + |\text{sh}(\Omega_+) \cap \text{sh}(\Omega_-)| \\
&= |\text{sh}(\Omega_+)| + |\text{sh}(\Omega_-)| \\
&\geq |\Omega_+| + |\Omega_-| = |\Omega|
\end{aligned}$$

which completes the induction as desired. □

We will primarily use the theorem as the following corollary, which says that if you have half of the points in terms of cardinality, you get half of the dimension.

Corollary 2.3.10. *If* $|\Omega| \geq 2^{n-1}$ *then there exists* $\sigma \subseteq \{1, \ldots, n\}$ *with* $|\sigma| \geq \lceil \frac{n+1}{2} \rceil \geq \frac{n}{2}$ *such that* $\text{Proj}_{\mathbb{R}^\sigma} \Omega = \{\pm 1\}^\sigma$.

2-15-16: Last time we left off with the proof of the Sauer-Shelah lemma. To remind you, we were finding ways to find interesting subsets where matrices behave well. Now recall we had a linear algebraic fact which I owe you; I will prove it in an analytic way. The proof has been moved to Section 2.

4 Proof of RIP

4.1 Step 1

Now we need another geometric lemma for the proof of Theorem 2.1.7, the restricted invertibility principle.

Lemma 2.4.1 (Step 1)**.** lem:step1 *Fix $m, n, r \in \mathbb{N}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator with rank$(A) \geq r$. For every $\tau \subseteq \{1, \ldots, m\}$, denote*

$$E_\tau = (\text{span}((Ae_j)_{j\in\tau}))^\perp .$$

Then there exists $\tau \subseteq \{1, \ldots, m\}$ with $|\tau| = r$ such that for all $j \in \tau$,

$$\|\text{Proj}_{E_{\tau\setminus\{j\}}} Ae_j\|_2 \geq \frac{1}{\sqrt{m}} \left(\sum_{i=1}^m s_i(A)^2\right)^{1/2} .$$

Basically we're taking the projection of the j^{th} column onto the orthogonal completement of the span of the subspace of all columns in the set except for the j^{th} one, and bounding the norm of that by a dimension term and the square root of the sum of the eigenvalues. (This is sharp asymptotically, and may in fact even be sharp as written too—I need to check. Check this?)

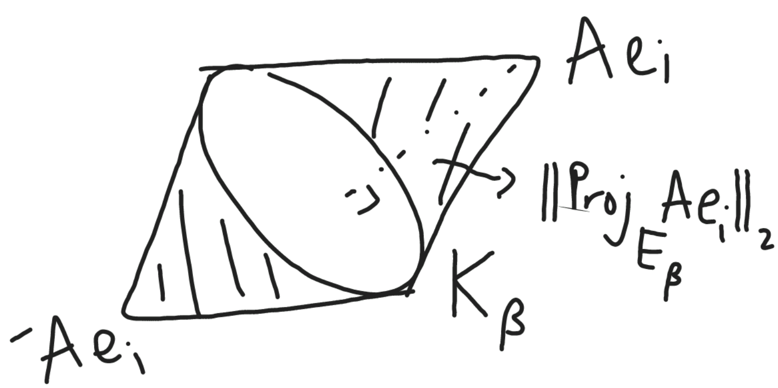
Proof. For every $\tau \subseteq \{1, \ldots, m\}$, denote

$$K_\tau = \text{conv}\left(\{\pm Ae_j\}_{j\in\tau}\right)$$

The idea is to make the convex hull have big volume. Once we do that, wewill get all these inequalities for free. Let $\tau \subseteq \{1, \ldots, m\}$ be the subset of size r that maximizes $\text{vol}_r(K_\tau)$. We know that $\text{vol}_r(K_\tau) > 0$. Observe that for any $\beta \subseteq \{1, \ldots, m\}$ of size $r - 1$ and $i \notin \beta$, we have

$$K_{\beta\cup\{i\}} = \text{conv}\left(K_\beta \cup \{\pm Ae_i\}\right),$$

which is a double cone.



What is the height of this cone? It is $\|\text{Proj}_{E_\beta} Ae_i\|_2$, as E_β is the orthogonal complement of the space spanned by β. Therefore, the r-dimensional volume is given by

$$\text{vol}_r(K_{\beta\cup\{i\}}) = 2 \cdot \frac{\text{vol}_{r-1}(K_\beta) \cdot \|\text{Proj}_{E_\beta} Ae_i\|_2}{r}$$

Because $|\tau| = r$ is the maximizing subset of F_Ω, for any $j \in \tau$ and $i \in \{1, \dots, m\}$, choosing $\beta = \tau \setminus \{j\}$, we get

$$\begin{aligned} &\mathrm{vol}_r(K_{\beta\cup\{j\}}) \geq \mathrm{vol}_r(K_{\beta\cup\{i\}}) \\ \implies &\|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} Ae_j\|_2 \geq \|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} Ae_i\|_2. \end{aligned}$$

for every $j \in \tau$ and $i \in \{1, \dots, m\}$. Summing,

$$m\|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} Ae_j\|_2^2 \geq \sum_{i=1}^m \|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} Ae_i\|_2^2 = \|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} A\|_{S_2}^2.$$

Then, for all $j \in \tau$,

eq:rip-s1-1
$$\|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} Ae_j\|_2 \geq \frac{1}{\sqrt{m}}\|\mathrm{Proj}_{E_{\tau\setminus\{j\}}} A\|_{S_2} \tag{2.12}$$

Let's denote $P = \mathrm{Proj}_{E_{\tau\setminus\{j\}}}$. Note P is an orthogonal projection of rank $r-1$. Then,

eq:rip-s1-2
$$\begin{aligned} \|PA\|_{S_2}^2 &= \mathrm{Tr}((PA)^*(PA)) = \mathrm{Tr}(A^*P^*PA) = \mathrm{Tr}(A^*PA) = \mathrm{Tr}(AA^*P) \\ &= \mathrm{Tr}(AA^*) - \mathrm{Tr}(AA^*(I-P)) \geq \sum_{i=1}^m s_i(A)^2 - \sum_{i=1}^{r-1} s_i(A)^2 = \sum_{i=r}^m s_i(A)^2 \end{aligned} \tag{2.13}$$

using the Ky Fan maximal principle 2.2.1, since $I - P$ is a projection of rank $m - r + 1$.

Putting (2.12) and (2.13) together gives the result. $\square$

4.2 Step 2

In our proof of the restricted invertibility principle, this is the first step. Before proving it, let me just tell you what the second step looks like.

Lemma 2.4.2 (Step 2). lem:step2 *Let $k, m, n \in \mathbb{N}$, $A : \mathbb{R}^m \to \mathbb{R}^n$, $\mathrm{rank}(A) > k$. Let $\omega \subseteq \{1, \dots, m\}$ with $|\omega| = \mathrm{rank}(A)$ such that $\{Ae_j\}_{j\in\omega}$ are linearly independent. Denote for every $j \in \Omega$*

$$F_j = E_{\omega\setminus\{j\}} = \left(\mathrm{span}(Ae_i)_{i\in\omega\setminus\{j\}}\right).$$

Then there exists $\sigma \subseteq \omega$ with $|\sigma| = k$ such that

$$\|\left(AJ_\sigma\right)^{-1}\|_{S_\infty} \lesssim \frac{\sqrt{\mathrm{rank}(A)}}{\sqrt{\mathrm{rank}(A) - k}} \cdot \max_{j\in\omega} \frac{1}{\|\mathrm{Proj}_{F_j} Ae_j\|}$$

Most of the work is in the second step. First we pass to a subset where we have some information about the shortest possible orthogonal project. But Step 1 saves us by bounding what this can be. Here we use the Grothendieck inequality, Sauer-Shelah, etc. Everything: It's simple, but it kills the restricted invertibility principle.

Proof of Theorem 2.1.7 given Step 1 and 2. Take $A : \mathbb{R}^m \to \mathbb{R}^n$. By Step 1 (Lemma 2.4.1), we can find subset $\tau \subseteq \{1, \ldots, m\}$ with $|\tau| = r$ such that for all $j \in \tau$,

$$\text{eq:step1} \|\text{Proj}_{E_{\tau \setminus \{j\}}} Ae_j\|_2 \geq \frac{1}{\sqrt{m}} \left(\sum_{i=r}^{m} s_i(A)^2 \right)^{1/2}. \tag{2.14}$$

Now we apply Step 2 (Lemma 2.4.2) to AJ_τ, using $\omega = \tau$, and find a further subset $\sigma \subseteq \tau$ such that

$$\begin{aligned} \| (AJ_\sigma)^{-1} \|_{S_\infty} &\leq \min_{k<r<\text{rank}(A)} \sqrt{\frac{\text{rank}(A)}{\text{rank}(A) - r}} \max_{j \in \omega} \frac{1}{\left\| \text{Proj}_{F_j} Ae_j \right\|} \\ &\leq \min_{k<r<\text{rank}(A)} \sqrt{\frac{mr}{(r-k)\sum_{i=r}^{m} s_i(A)^2}} \end{aligned}$$

which we get by plugging directly in r for the rank and using Step 1 (2.14) to get the denominator. □

2-17

Now we prove Step 2 (Lemma 2.4.2). Note we can assume $\omega = \{1, \ldots, m\}$ and that the rank is m.

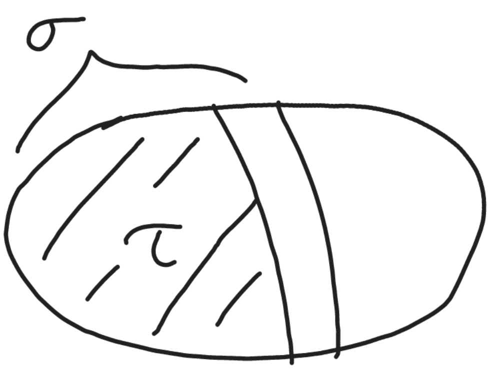
First we need some lemmas.

Lemma 2.4.3. lem:rip-step2-1 *Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be such that $\{Ae_j\}_{j=1}^m$ are linearly independent, and $\sigma \subseteq \{1, \ldots, m\}$, $t \in \mathbb{N}$. Then there exists $\tau \subseteq \sigma$ with*

$$|\tau| \geq \left(1 - \frac{1}{2^t}\right) |\sigma|,$$

such that denoting $\theta = \tau \cup (\{1, \ldots, m\} \setminus \sigma)$, $F_j = (\text{span}(\{Ae_j\}_{i \neq j}))^\perp$, and $M = \max_{j \in \omega} \frac{1}{\|\text{Proj}_{F_j} Ae_j\|}$, we have that for all $\sigma \in \mathbb{R}^\theta$,

$$\sum_{i \in \tau} |a_i| \leq 2^{\frac{t}{2}} M \sqrt{|\sigma|} \left\| \sum_{i \in \theta} a_i Ae_i \right\|_2.$$



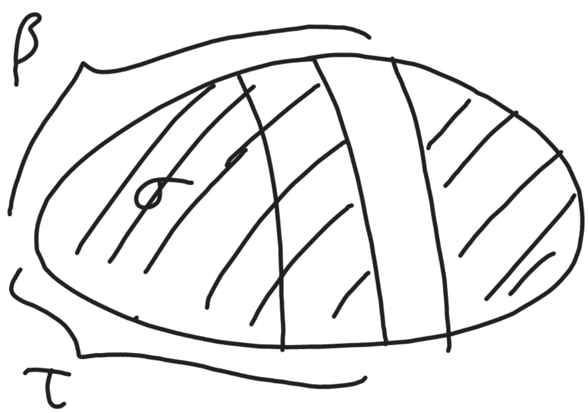
This is proved by a nice inductive argument.

Proof. TODO next time

∎

Lemma 2.4.4. lem:rip-step2-2 *Let $m, n, t \in \mathbb{N}$ and $\beta \subseteq \{1, \dots, m\}$. Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator such that $\{Ae_j\}_{j=1}^m$ are linearly independent. Then there exist two subsets $\sigma \subseteq \tau \subseteq \beta$ such that $|\tau| \geq \left(1 - \frac{1}{2^t}\right)|\beta|$, $|\tau \backslash \sigma| \leq \frac{|\beta|}{4}$, and if we denote $\theta = \tau \cup (\{1, \dots, m\} \backslash \beta)$, $M = \max_{j \in \omega} \frac{1}{\|\mathrm{Proj}_{F_j} Ae_j\|}$, then*

$$\left\| \mathrm{Proj}_{\mathbb{R}^\sigma} (AJ_\theta)^{-1} \right\|_{S_\infty} \lesssim 2^{\frac{t}{2}} M.$$



Proof of Lemma 2.4.4 from Lemma 2.4.3. Apply Lemma 2.4.3 with $\sigma = \beta$. Basically, we're going to inductively construct a subset $\sigma = \beta$ of $\{1, \cdots, m\}$ onto which to project, so that we can control the l_1 norm of the subset τ by the l_2 norm on a slightly greater set θ via Lemma 2.4.3.

We find $\tau \subseteq \beta$ with $|\tau| \geq \left(1 - \frac{1}{2^t}\right)|\beta|$ such that

$$\sum_{i \in \tau} |a_i| \leq 2^{\frac{t}{2}} M \sqrt{|\beta|} \left\| \sum_{i \in \theta} a_i A e_i \right\|_2 .$$

Rewriting gives that

$$\begin{aligned} \forall a \in \mathbb{R}^\sigma, \qquad & \|\mathrm{Proj}_{\mathbb{R}^\tau} a\| \leq 2^{\frac{t}{2}} M \sqrt{|\beta|}\, \|AJ_\theta a\|_2 \\ \implies & \left\| \mathrm{Proj}_{\mathbb{R}^\tau} (AJ_\theta)^{-1} \right\|_{\ell_2^\theta \to \ell_1^\tau} \lesssim 2^{\frac{t}{2}} M \sqrt{|\beta|}. \end{aligned}$$

Denote $\varepsilon = \frac{|\beta|}{4|\tau|}$. By Lemma 2.3.7, there exists $\sigma \subseteq \tau$, $|\sigma| \geq (1 - \varepsilon)|\tau|$ such that

$$\begin{aligned} \left\| \mathrm{Proj}_{\mathbb{R}^\sigma} (AJ_\theta)^{-1} \right\|_{S^\infty} = \left\| \mathrm{Proj}_{\mathbb{R}^\sigma} \mathrm{Proj}_{\mathbb{R}^\tau} (AJ_\theta)^{-1} \right\|_{S^\infty} &\lesssim \sqrt{\frac{\pi}{2\varepsilon|\tau|}} 2^{\frac{t}{2}} M \sqrt{|\beta|} \\ &\lesssim 2^{\frac{t}{2}} M. \end{aligned}$$

∎

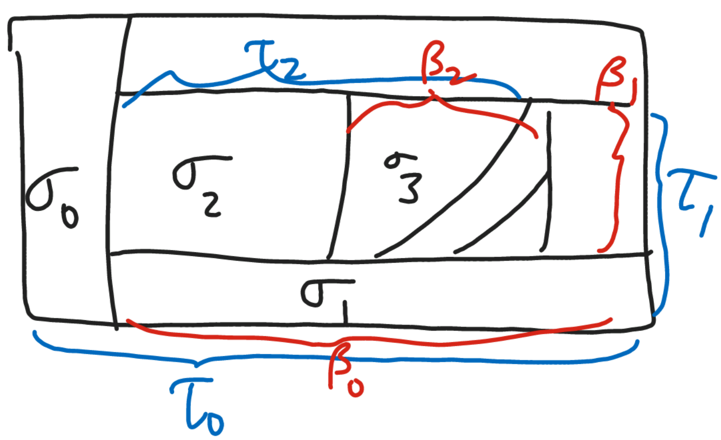
Now, we have basically finished making use of Lemma 2.4.3 in the proof of Lemma 2.4.4. It remains to use Lemma 2.4.4 to prove the full second step in our proof of the Restricted Invertibility Principle.

Proof of Lemma 2.4.2 (Step 2). Fix an integer r such that $\frac{1}{2^{2r+1}} \leq 1 - \frac{k}{m} \leq \frac{1}{2^{2r}}$. Proceed inductively as follows. First set

$$\begin{aligned} \tau_0 &= \{1, \ldots, m\} \\ \sigma_0 &= \phi. \end{aligned}$$

Suppose $u \in \{0, \ldots, r+1\}$ and we constructed $\sigma_k, \tau_k \subseteq \{1, \ldots, m\}$ such that if we denote $\beta_u = \tau_u \backslash \sigma_u$, $\theta_u = \tau_u \cup (\{1, \ldots, m\} \backslash \beta_{u-1})$, then

1. $\sigma_u \subseteq \tau_u \subseteq \beta_{u-1}$
2. $|\tau_u| \geq \left(1 - \frac{1}{2^{2r-u+4}}\right) |\beta_{u-1}|$
3. $|\beta_u| \leq \frac{1}{4} |\beta_{u-1}|$
4. $\|\text{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\theta_u})^{-1}\|_{S_\infty} \lesssim 2^{r - \frac{u}{2}} M$.



Let $H = 2r - u + 4$. For instance, $|\tau_1| \geq \left(1 - \frac{1}{2^{2r+3}}\right) |\beta_0|$. What is the new β?

For the inductive step, apply Lemma 2.4.4 on β_{u-1} with $t = 2r - u + 4$ to get $\sigma_u \subseteq \tau_u \subseteq \beta_{u-1}$ such that $|\tau_u| \geq \left(1 - \frac{1}{2^{2r-u+4}}\right) |\beta_{u-1}|$, $|\tau_u \backslash \sigma_u| \leq \frac{|\beta_{u-1}|}{4}$ As we induct, we are essentially building up more σ_u to eventually produce the invertible subset over which we will project. Note that the size of the θ set is decreasing as we proceed inductively.

eq:rip-s2-1
$$\left\| \text{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\theta_u})^{-1} \right\| \lesssim 2^{r - u/2} M. \tag{2.15}$$

We know $|\beta_{u-1}| \le \frac{m}{4^{u-1}}$,

$$\begin{aligned}
\beta_{u-1} &= \beta_u \sqcup \sigma_u \sqcup (\beta_{u-1} \backslash \tau_u) \\
|\beta_{u-1}| &= |\beta_u| + |\sigma_u| + (|\beta_{u-1}| - |\tau_u|) \\
|\sigma_u| &= |\beta_{u-1}| - |\beta_u| - (|\beta_{u-1}| - |\tau_u|) \\
&\ge |\beta_{u-1}| - |\beta_u| - \frac{|\beta_{u-1}|}{2^{2r-u+4}} \\
&\ge |\beta_{u-1}| - |\beta_u| - \frac{m}{2^{2r+u+2}}.
\end{aligned}$$

Our choice for the invertible subset is

$$\sigma = \bigsqcup_{u=1}^{r+1} \sigma_u.$$

Telescoping gives

$$\begin{aligned}
|\sigma| = \sum_{u=1}^{r+1} |\sigma_u| &\ge |\beta_0| - |\beta_{r+1}| - \frac{m}{2^{2r+2}} \sum_{u=1}^{\infty} \frac{1}{2^u} \\
&\ge m - \frac{m}{4^{r+1}} - \frac{m}{2^{2r+2}} \\
&= m\left(1 - \frac{1}{2^{2r+1}}\right) \ge m\frac{k}{m} = k.
\end{aligned}$$

Observe that $\sigma \subseteq \bigcap_{u=1}^{r+1} \theta_u$ and for every u,

$$\begin{aligned}
&\sigma_u, \ldots, \sigma_{r+1} \subseteq \tau_u \\
&\sigma_1, \ldots, \sigma_{u-1} \subseteq \{1, \ldots, m\} \backslash \beta_{u-1}
\end{aligned}$$

This allows us to use the conclusion for all the σ_u's at once.

For $a \in \mathbb{R}^\sigma$, $J_\sigma a \subseteq J_{\theta_u} \mathbb{R}^{\theta_u}$,

$$\mathrm{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\theta_u})^{-1} (AJ_\sigma) a = \mathrm{Proj}_{\mathbb{R}^{\sigma_u}} J_\sigma a.$$

since $A^{-1}A = I$ and projecting a subset onto its containing set is just the subset itself. Then, breaking $J_\sigma a$ into orthogonal components,

$$\begin{aligned}
\|J_\sigma a\|_2^2 &= \sum_{u=1}^{r+1} \|\mathrm{Proj}_{\mathbb{R}^{\sigma_u}} J_\sigma a\|_2^2 \\
&= \sum_{u=1}^{r+1} \left\|\mathrm{Proj}_{\mathbb{R}^{\sigma_u}} (AJ_{\sigma_u})^{-1} (AJ_\sigma) a\right\|_2^2 \\
&\lesssim \sum_{u=1}^{r+1} 2^{2r-u} M^2 \|AJ_\sigma a\|_2^2 && \text{by (2.15)} \\
&\lesssim 2^{2r} M^2 \|AJ_\sigma a\|_2^2
\end{aligned}$$

$$\leq \frac{M^2}{1-\frac{k}{m}} \|AJ_\sigma a\|_2^2$$
$$\|J_\sigma a\|_2 \leq \sqrt{\frac{m}{m-k}} M \|AJ_\sigma a\|_2$$

Since this is true for all $a \in \mathbb{R}^\sigma$,

$$\implies \left\|(AJ_\sigma)^{-1}\right\|_{S_\infty} \lesssim \sqrt{\frac{m}{m-k}} M.$$

□

An important aspect of this whole proof to realize is the way we took a first subset to give one bound, and then took another subset in order to apply the Little Grothendieck inequality. In other words, we first took a subset where we could control l_1 to l_2, and then took a further subset to get the operator norm bounds.

Also, notice that in this approach, we construct our "optimal subset" of the space inductively: However, doing this in general is inefficient. It would be nice to construct the set greedily instead for algorithmic reasons, if we wanted to come up with a constructive proof. I have a feeling that if we were to avoid this kind of induction, it would involve not using Sauer-Shelah at all.

How can we make this theorem algorithmic?

The way the Pietsch Domination Theorem 2.3.5 worked was by duality. We look at a certain explicitly defined convex set. We found a separating hyperplane which must be a probability measure. Then we had a probabilistic construction. This part is fine.

The bottleneck for making this an algorithm (I do believe this will become an algorithm) consists of 2 parts:

1. Sauer-Shelah lemma 2.3.9: We have some cloud of points in the boolean cube, and we know there is some large subset of coordinates (half of them) such that when you project to it you see the full cube. I'm quite certain that it's NP-hard in general. (Work is necessary to formulate the correct algorithmic Sauer-Shelah lemma. How is the set given to you?). In fact, formulating the right question for an algorithmi Sauer-Shelah is the biggest difficulty.

   We only need to answer the algorithmic question tailored to our sets, which have a simple description: the intersection of an ellipsoid with a cube. There is a good chance that there is a polynomial time algorithm in this case. This question has other applications as well (perhaps one could generalize to the intersection of the cube with other convex bodies). [3]

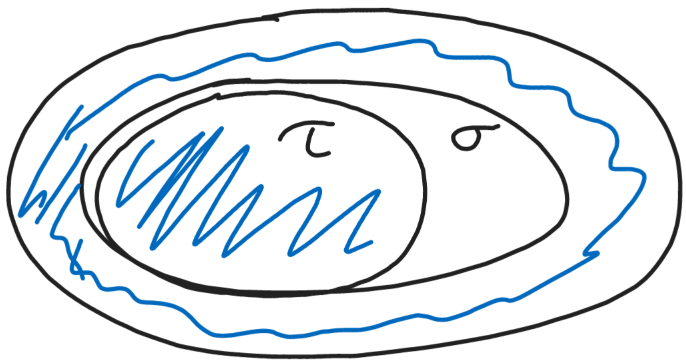
2. The second bottleneck is finding a subset with maximal volume.

   It's another place where we chose a subset, the subset that maximizes the volume out of all subsets of a given size (Lemma 2.4.1). Specifically, we want the set of columns

[3] There could be an algorithm out there, but I couldn't find one in the literature.

of a matrix that maximizes the volume of the convex hull. Computing the volume of the convex hull requires some thought. Also, there are $\binom{n}{r}$ subsets of size r; if $r = n/2$ there are exponentially many. We need a way to find subsets with maximum volume fast. There might be a replacement algorithm which approximately maximizes this volume.

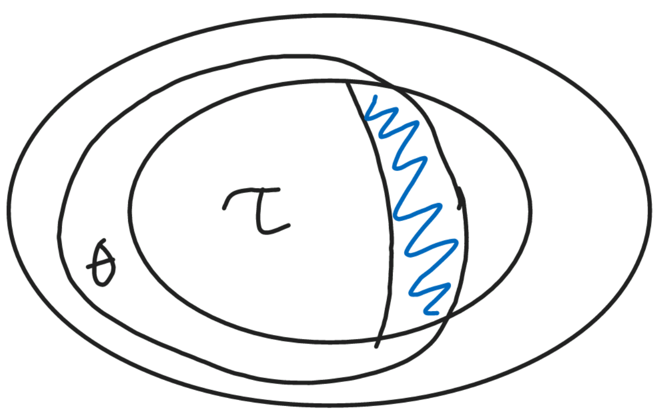
2-22: We are at the final lemma, Lemma 2.4.3 in the Restricted Invertiblity Theorem.



The proof of this is an inductive application of the Sauer-Shelah lemma. A very important idea comes from Giannopoulos. If you naively try to use Sauer-Shelah, it won't work out. We will give a stronger statement of the previous lemma which we can prove by induction.

Lemma 2.4.5 (Stronger version of Lemma 2.4.3)**.** lem:SS-induct-stronger *Take $m, n \in \mathbb{N}$, $A : \mathbb{R}^m \to \mathbb{R}^n$ a linear operator such that $\{Ae_j\}_{j=1}^m$ are linearly independent. Suppose that $k \ge 0$ is an integer and $\sigma \subseteq \{1, \dots, m\}$. Then there exists $\tau \subseteq \sigma$ with $|\tau| \ge (1 - \frac{1}{2^k})|\sigma|$ such that for every $\theta \supseteq \tau$ for all $a \in \mathbb{R}^m$ we have*

$$\text{eq:rip21-1} \sum_{i\in\tau} |a_i| \le M\sqrt{|\sigma|}\left(\sum_{r=1}^{k} 2^{r/2}\right)\left\|\sum_{i\in\theta} a_i Ae_i\right\|_2 + (2^k - 1)\sum_{i\in\theta\cap(\sigma\setminus\tau)} |a_i| \tag{2.16}$$



Lemma 2.4.3 (all we need to complete the proof of the Restricted Invertibility Principle) is the case where $\theta = \tau \cup \{1, \cdots, m\} \setminus \sigma$, $t = k$.

We prove the stronger version via induction on k.

Proof. As k becomes bigger, we're allowed to take more of the set σ. The idea is to use Sauer-Shelah, taking half of σ, and then half of what remains at each step.

For $k = 0$, the statement is vacuous, because we can take τ to be the empty set. By induction, assume that the statement holds for k: we have found $\tau \subseteq \sigma$ such that $|\tau| \geq (1 - \frac{1}{2^k})|\sigma|$ and (2.16) holds for every $\tau \subseteq \theta$. If $\sigma = \tau$ already, then τ satisfies (2.16) for $k+1$ as well, so WLOG $|\sigma \setminus \tau| > 0$. Now define v_j is the projection

$$v_j = \frac{\mathrm{Proj}_{F_j} Ae_j}{\left\|\mathrm{Proj}_{F_j} Ae_j\right\|_2^2}$$

Then $\langle v_i, Ae_j\rangle = \delta_{ij}$ by definition since (v_i) is a dual basis for the Ae_j's.

Now we want to user Sauer-Shelah so we're going to define a certain subset of the cube. Define

$$\Omega = \left\{\epsilon \in \{\pm 1\}^{\sigma\setminus\tau} : \left\|\sum_{i\in\sigma\setminus\tau} \epsilon_i v_i\right\|_2 \leq M\sqrt{2|\sigma \setminus \tau|}\right\}$$

which is an ellipsoid intersected with the cube (it is not a sphere since the v_is are not orthogonal). Then we have

$$\begin{aligned} M^2|\sigma \setminus \tau| &\geq \sum_{i\in\sigma\setminus\tau} \frac{1}{\|\mathrm{Proj}_{F_j} Ae_j\|^2} = \sum_{j\in\sigma\setminus\tau} \|v_j\|_2^2 \\ &= \frac{1}{2^{|\sigma\setminus\tau|}} \sum_{\epsilon\in\{\pm1\}^{\sigma\setminus\tau}} \left\|\sum_{j\in\sigma\setminus\tau} \epsilon_j v_j\right\|_2^2. \end{aligned}$$

where the last step is true for any vectors (sum the squares and the pairwise correlations disappear).

Using Markov's inequality this is

$$\geq \frac{1}{2^{|\sigma\setminus\tau|}} \left(2^{|\sigma\setminus\tau|} - |\Omega|\right) M^2 2|\sigma \setminus \tau|$$

which gives

$$|\Omega| > 2^{|\sigma\setminus\tau|-1}$$

Then by Sauer-Shelah (Lemma 2.3.9, there exists $\beta \subseteq \sigma \setminus \tau$ such that

$$\mathrm{Proj}_{\mathbb{R}^\beta} \Omega = \{\pm 1\}^\beta$$

and

$$|\beta| \geq \frac{1}{2}|\sigma \setminus \tau|$$

Define $\tau^* = \tau \cup \beta$. We will show that τ^* satisfies the inductive hypothesis with $k+1$. Each time we find a certain set of coordinates to add to what we have before. $|\tau^*|$ is the correct size because

$$|\tau^*| = |\tau| + |\beta| \geq |\tau| + \frac{|\sigma| - |\tau|}{2} = \frac{|\tau| + |\sigma|}{2} \geq \left(1 - \frac{1}{2^{k+1}}\right) |\sigma|$$

where we used that $|\tau| \geq \left(1 - \frac{1}{2^k}\right)|\sigma|$. So at least τ^* is the right size.

Now, suppose $\theta \supseteq \tau^*$. For every $a \in \mathbb{R}^m$, we claim there exists some $\epsilon \in \Omega$ such that $\forall j \in \beta$ such that $\epsilon_j = \text{sign}(a_j)$. For any β, we can find some vector in the cube that has the sign pattern of our given vector a. What does being in Ω mean? It means that at least the dual basis is small there. $\epsilon \in \Omega$ says that

$$\text{eq:step21-2} \| \sum_{i \in \sigma \setminus \tau} \epsilon_i v_i \|_2 \leq M\sqrt{2|\sigma \setminus \tau|} \leq \frac{M\sqrt{2|\sigma|}}{2^{k/2}} \tag{2.17}$$

That was how we chose our ellipsoid.

$$\sum_{i \in \beta} |a_i| = \left\langle \sum_{i \in \beta} a_i A e_i, \sum_{i \in \sigma \setminus \tau} \epsilon v_i \right\rangle$$

because the v_i's are a dual basis so $\langle Ae_i, v_j \rangle = \delta_{ij}$, and $\beta \subseteq \sigma \backslash \tau$. We only know the ϵ_i are the signs when you're inside β. This equals

$$= \left\langle \sum_{i \in \theta} a_i A e_i, \sum_{i \in \sigma \setminus \tau} \epsilon v_i \right\rangle - \sum_{i \in (\theta \setminus \beta) \cap (\sigma \setminus \tau)} \epsilon_i a_i$$

Note that $(\theta \setminus \beta) \cap (\sigma \setminus \tau) = \theta \cap (\sigma \setminus \tau^*)$. In this set we can't control the signs ϵ_i. By Cauchy-Schwarz and (2.17), this is

$$\begin{aligned} &\leq \left\| \sum_{i \in \theta} a_i A e_i \right\|_2 \cdot \left\| \sum_{i \in \sigma \setminus \tau} \epsilon v_i \right\|_2 + \sum_{i \in \theta \cap (\sigma \setminus \tau^*)} |a_i| \\ &\leq \left\| \sum_{i \in \theta} a_i A e_i \right\|_2 \cdot \frac{M\sqrt{2|\sigma|}}{2^{k/2}} + \sum_{i \in \theta \cap (\sigma \setminus \tau^*)} |a_i|. \end{aligned}$$

Because the conclusion of Sauer-Shelah told us nothing about the signs of ϵ_i outside β, so we just take the worst possible thing.

Summarizing,

$$\sum_{i \in \beta} |a_i| \leq \frac{M\sqrt{2|\sigma|}}{2^{k/2}} \left\| \sum_{i \in \theta} a_i A e_i \right\|_2 + \sum_{i \in \theta \setminus (\sigma \setminus \tau^*)} |a_i|$$

Using the inductive step,

$$\begin{aligned}
\sum_{i\in\tau^*}|a_i| &= \sum_{i\in\tau}|a_i| + \sum_{i\in\beta}|a_i| \\
&\le M\sqrt{|\sigma|}\alpha_k \left\|\sum_{i\in\theta} a_i A e_i\right\|_2 + \left(2^k-1\right)\sum_{i\in\theta\cap(\sigma\setminus\tau)}|a_i| + \sum_{i\in\beta}|a_i| \\
&= \alpha_k\sqrt{|\sigma|}\left\|\sum_{i\in\theta} a_i A e_i\right\|_2 + \left(2^k-1\right)\sum_{i\in\theta\cap(\sigma\setminus\tau^*)}|a_i| + 2^k\sum_{i\in\beta}|a_i|
\end{aligned}$$

In the last step, we moved β from the second to the third term. Now use what we got before for the bound on $\sum_{i\in\beta}|a_i|$ and plug it in to get

$$\le \left(\alpha_k + 2^{(k+1)/2}\right)\sqrt{|\sigma|}\left\|\sum_{i\in\theta} a_i A e_i\right\|_2 + \left(2^{k+1}-1\right)\sum_{i\in\theta\cap(\sigma\setminus\tau^*)}|a_i|$$

which is exactly the inductive hypothesis.[4] □

Remark 2.4.6 (Algorithmic Sauer-Shelah)**:** We only used Sauer-Shelah 2.3.9 for intersecting cubes with an ellipsoid, so to make RIP algorithmic, we need an algorithm for finding these particular intersections. Moreover, the ellipsoids are big is because we ask that the 2-norm is at most $\sqrt{2}$ times the expectation. An efficient algorithm probably exists.

Afterwards, wecould ask for higher dimensional shapes. I've seen some references that worked for Sauer-Shelah when sets were of a special form, namely of size $o(n)$. This is something more geometric. I don't think there's literature about Sauer-Shelah for intersection of surfaces with small degree. This is a tiny motivation to do it, but it's still interesting independently.

[4] I looked through the original Gianpopoulous paper, and it was clear he tried out many many things to find which inductive hypothesis makes everything go through cleanly. You want to bound an l_1 sum from above, so you want to use duality, and then use Sauer-Shelah to get signs such that the norm of the dual-basis is small.

Chapter 3

Bourgain's Discretization Theorem

1 Outline

We will prove Bourgain's Discretization Theorem. This will take maybe two weeks, and has many interesting ingredients along the way. By doing this, we will also prove Ribe's theorem, which is what we stated at the beginning.

Let's remind ourselves of the definition.

Definition (Definition 1.2.3, discretization modulus)**:** $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ are Banach spaces. Let $\epsilon \in (0,1)$. Then $\delta_{X \hookrightarrow Y}(\epsilon)$ is the supremum over $\delta > 0$ such that for every δ-net $\mathcal{N}_\delta$ of the unit ball of X, the distortion

$$C_Y(X) \le \frac{C_Y(\mathcal{N}_\delta)}{1-\epsilon}$$

C_Y is smallest bi-Lipschitz distortion by which you can embed X into Y. There are ideas required to get $1-\epsilon$. For example, for $\epsilon = \frac{1}{2}$, the equation says is that if we succeed in embedding a δ-net into Y, then we succeeded in the full space with a distortion twice as much. A priori it's not even clear that there exists such a δ. There's a nontrivial compactness argument needed to prove its existence. We will just prove bounds on it assuming it exists.

Now, Bourgain's discretization theorem says

Theorem (Bourgain's discretization theorem 1.2.4)**.** *If dim*$(X) = n$*, dim*$(Y) = \infty$*, then*

$$\delta_{X \hookrightarrow Y}(\epsilon) \ge e^{-\left(\frac{n}{\epsilon}\right)^{Cn}}$$

for C *a universal constant.*

Remark 3.1.1: It doesn't matter what the maps are for the $\mathcal{N}_\delta$, the proof will give a linear mapping and we won't end up needing them. Assuming linearity in the definition is not necessary.

Rademacher's theorem says that any mapping $\mathbb{R}^n \to \mathbb{R}^n$ is differentiable almost everywhere with bi-Lipschitz derivative. Yu can extend this to $n = \infty$, but you need some additional properties: Y must be the dual space, and the limit needs to be in the weak-$*$ topology. In this topology there exists a sequence of norm-1 vectors that tend to 0. Almost everywhere, this doesn't happen. The Principle of Local Reflexivity says Y^{**} and Y are not the same for infinite dimensional Y. The double dual of all sequences which tend to 0 is L_∞, a bigger space, but the difference between these never appears in finite dimensional phenomena.

1.1 Reduction to smooth norm

From now on, $B_X = \{x \in X : \|X\|_X \leq 1\}$ denotes the ball, and $S_X = \partial B = \{x \in X : \|X\|_X = 1\}$ denotes the boundary.

Later on we will be differentiating things without thinking about it, so we first prove that we can assume $\|\cdot\|_X$ is smooth on $X \setminus \{0\}$.

Lemma 3.1.2. *For all $\delta \in (0,1)$ there exists some δ-net of S_X with $|\mathcal{N}_\delta| \leq \left(1 + \frac{2}{\delta}\right)^n$.*

Proof. Let $\mathcal{N}_\delta \subseteq S_X$ be maximal with respect to inclusion such that $\|x-y\|_X > \delta$ for every distinct $x, y \in \mathcal{N}_\delta$. Maximality means that if $z \in S_X$, there exists $x \in \mathcal{N}_\delta$, $\|x-y\|_X \leq \delta$ (otherwise $N_\delta \cup \{z\}$ is a bigger δ-net). The balls $\{x + \frac{\delta}{2} B_X\}_{x \in \mathcal{N}_\delta}$ are pairwise disjoint and contained in $1 + \frac{\delta}{2} B_X$. Thus

$$\mathrm{vol}\left(\left(1 + \frac{\delta}{2}\right) B_X\right) \geq \sum_{X \in M_\delta} \mathrm{vol}\left(x + \frac{\delta}{2} B_X\right)$$

$$\left(1 + \frac{\delta}{2}\right)^n \mathrm{vol}(B_X) = |\mathcal{N}_\delta| \left(\frac{\delta}{2}\right)^n \mathrm{vol}(B_X)$$

[1] □

Reduction to smooth norm. Let $\mathcal{N}_\delta$ be a δ-net of S_X with $|\mathcal{N}_\delta| = N \leq (1 + \frac{2}{\delta})^n$. Given $z \in \mathcal{N}_\delta$, by Hahn-Banach we can choose z^* such that

1. $\langle z, z^* \rangle = 1$
2. $\|z^*\|_{X^*} = 1$.

Let k be an integer such that $N^{1/(2k)} \leq 1 + \delta$. Then define

$$\|x\| := \left(\sum_{z \in \mathcal{N}_\delta} \langle z^*, x \rangle^{2k}\right)^{1/(2k)}$$

Each term satisfies $|\langle z^*, x \rangle| \leq \|x\|_X$, so we know

$$\|x\| \leq N^{1/2k} \|x\|_X \leq (1+\delta)\|x\|_X.$$

[1] There are non-sharp bounds for the smallest size of a δ-net. There is a lot of literature about the relations between these things, but we just need an upper bound.

For the lower bound, if $x \in S_X$, then choose $z \in \mathcal{N}_\delta$ such that $\|x - z\|_X \le \delta$. Then $1 - \langle z^*, x\rangle = \langle z^*, z - x\rangle \le \|z - x\| \le \delta$, so $\langle z^*, x\rangle \ge 1 - \delta$, giving

$$\|x\| \ge (1-\delta)\,\|x\|_X\,.$$

Thus any norm is up to $1+\delta$ some equal to a smooth norm. δ was arbitrary, so without loss of generality we can assume in the proof of Bourgain embedding that the norm is smooth. □

The strategy to prove Bourgain's discretization theorem is as follows. Given a δ-net $\mathcal{N}_\delta \subseteq B_X$ with $\delta \le e^{-(n/\epsilon)^{Cn}}$, and we know $\exists\, f : \mathcal{N}_\delta \to Y$ such that $\frac{1}{D}\|x-y\|_X \le \|f(x) - f(y)\|_Y \le \|x - y\|_X$, i.e., we can embed with distortion D. Our goal is to show that if $\delta \le e^{-D^{Cn}}$ then this implies that there exists $T : X \to Y$ linear operator invertible $\|T\|\,\|T^{-1}\| \lesssim D$.

The steps are as follows.

1. Find the correct coordinate system (John ellipsoid), which will give us a dot product structure and a natural Laplacian. (We will need a little background in convex geometry.)

2. A priori f is defined on the net. Extend f to the whole space in a nice way (Bourgain's extension theorem) which doesn't coincide with the function on the net, but is not too far away from it.

3. Solve the Laplace equation. Start with some initial condition, and then evolve f according to the Poisson semigroup. This extended function is smooth the instant it flows a little bit away from the discrete function.

4. There is a point where the derivative satisfies what we want: The point exists by a pigeonhole style argument, but we won't be able to give it explicitly. This comes from estimates of the Poisson kernel. We will use Fourier analysis.

2-24

2 Bourgain's almost extension theorem

Theorem 3.2.1 (Bourgain's almost extension theorem)**.** *Let X be a n-dimensional normed space, Y a Banach space, $\mathcal{N}_\delta \subseteq S_X$ a δ-net of S_X, $\tau \ge C\delta$. Suppose $f : \mathcal{N}_\delta \to Y$ is L-Lipschitz. Then there exists $F : X \to Y$ such that*

1. $\|F\|_{Lip} \lesssim (1 + \frac{\delta n}{\tau})L$.
2. $\|F(x) - f(x)\|_Y \le \tau L$ *for all* $x \in \mathcal{N}_\delta$.
3. $\mathrm{Supp}(F) \subseteq (2+\tau)B_X$.
4. F *is smooth.*

Parts 3 and 4 will come "for free."

2.1 Lipschitz extension problem

Theorem 3.2.2 (Johnson-Lindenstrauss-Schechtman, 1986)**.** *Let X be a n-dimensional normed space $A \subseteq X$, Y be a Banach space, and $f : A \to Y$ be L-Lipschitz. There exists $F : X \to Y$ such that $F|_A = f$ and $\|F\|_{Lip} \lesssim nL$.*

We know a lower bound of $\sqrt{n}$; losing $\sqrt{n}$ is sometimes needed. (The lower bound for nets on the whole space is $\sqrt[4]{n}$.) A big open problem is what the true bound is.

This was done a year before Bourgain. Why didn't he use this theorem? This theorem is not sufficient because the Lipschitz constant grows with n.

We want to show that $\|T\| \|T^{-1}\| \lesssim D$. We can't bound the norm of T^{-1} with anything less than the distortion D, so to prove Bourgain embedding we can't lose anything in the Lipschitz constant of T; the Lipschitz constant can't go to ∞ as $n \to \infty$.

Bourgain had the idea of relaxing the requirement that the new function be strictly an extension (i.e., agree with the original function where it is defined). What's extremely important is that the new function be Lipschitz with a constant independent of n.

We need $\|F\|_{\text{Lip}} \lesssim L$. Let's normalize so $L = 1$.

When the parameter is $\tau = n\delta$, we want $\|f(x) - F(x)\| \lesssim n\delta$. Note δ is small (less than the inverse of any polynomial), so losing n is nothing.[2]

How sharp is Theorem 3.2.1? Given a 1-Lipschitz function on a δ-net, if we want to almost embed it without losing anything, how close can close can we guarantee it to be from the original function

Theorem 3.2.3. *There exists a n-dimensional normed space X, Banach space Y, $\delta > 0$, $\mathcal{N}_\delta \subseteq S_X$ a δ-net, 1-Lipschitz function $f : \mathcal{N}_\delta \to Y$ such that if $f : X \to Y$, $\|F\|_{Lip} \lesssim 1$ then there exists $x \in \mathcal{N}_\delta$ such that*

$$\|F(x) - f(x)\| \gtrsim \frac{n}{e^{c\sqrt{\ln n}}}.$$

Thus, what Bourgain proved is essentially sharp. This is a fun construction with Grothendieck's inequality.

Our strategy is as follows. Consider $P_t * F$, where

$$P_t(x) = \frac{C_n t}{(t^2 + \|x\|_2^2)^{\frac{n+1}{2}}}, \quad C_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}},$$

and P_t is the **Poisson kernel**. Let $(TF)(x) = (P_t * F)'(x)$; T is linear and $\|T\| \lesssim 1$. As $t \to 0$, this becomes closer to F. We hope when $t \to 0$ that it is invertible. This is true. We give a proof by contradiction (we won't actually find what t is) using the pigeonhole principle.

The Poisson kernel depends on an Euclidean norm in it. In order for the argument to work, we have to choose the right Euclidean norm.

[2]If people get δ to be polynomial, then we'll have to start caring about n.

2.2 John ellipsoid

Theorem 3.2.4 (John's theorem)**.** *Let X be a n-dimensional normed space, identified with $\mathbb{R}^n$. Then there exists a linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $\frac{1}{\sqrt{n}} \|Tx\|_2 \le \|x\|_X \le \|Tx\|_2$ for all $x \in X$.*

John's Theorem says we can always find T which sandwiches the X-norm up to a factor of $\sqrt{n}$. "Everything is a Hilbert space up to a $\sqrt{n}$ factor."

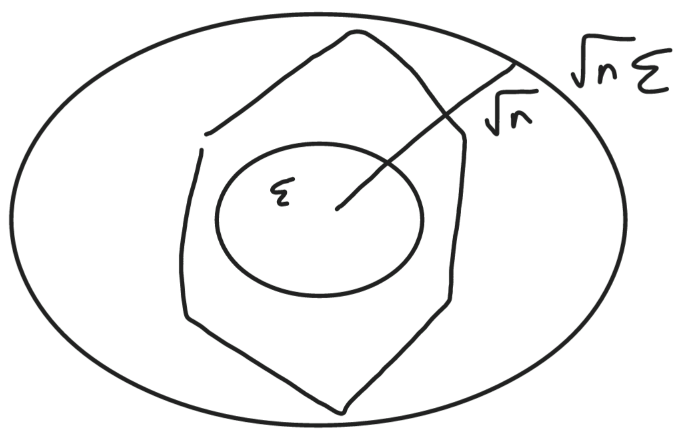
Proof. Without loss of generality $D \le n$. Let $\mathcal{E}$ be an ellipsoid of maximal volume contained in B_X.

$$\max\left\{\mathrm{vol}(SB_{\ell_2^n}) : S \in M_n(\mathbb{R}), SB_{\ell_2^n} \subseteq B_X\right\}.$$

$\mathcal{E}$ exists by compactness. Take $T = S^{-1}$.

The goal is to show that $\sqrt{n}\mathcal{E} \supseteq B_X$. We can choose the coordinate system such that $\mathcal{E} = B_{\ell_2^n} = B_2^n$.

Suppose by way of contradiction that $\sqrt{n}B_2^n \not\supseteq B_X$. Then there exists $x \in B_X$ such that $\|x\|_X \le 1$ yet $\|x\|_2 > \sqrt{n}$.



Denote $d = \|x\|_2, y = \frac{x}{d}$. Then $\|y\|_2 = 1$ and $dy \in B_X$.

By applying a rotation, we can assume WLOG $y = e_1$.

Claim: Define for $a > 1, b < 1$

$$\mathcal{E}_{a,b} := \left\{\sum_{i=1}^n t_i e_i : \left(\frac{t_1}{a}\right)^2 + \sum_{i=2}^n \left(\frac{t_i}{b}\right)^2 \le 1\right\}.$$

Stretching by a and squeezing by b can make the ellipse grow and stay inside the body, contradicting that it is the minimizer. More precisely, we show that there exists $a > 1$ and $b < 1$ such that $\mathcal{E}_{a,b} \subseteq B_X$ and $\mathrm{Vol}(\mathcal{E}_{a,b}) > \mathrm{Vol}(B_2^n)$. This is equivalent to $ab^{n-1} > 1$.

We need a lemma with some trigonometry.

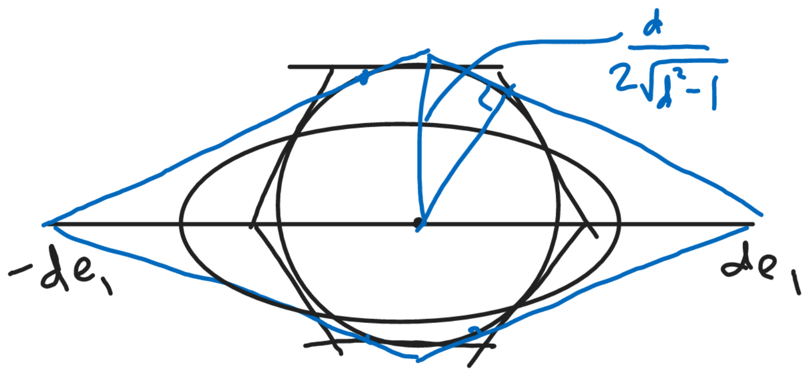
Lemma 3.2.5 (2-dimensional lemma)**.** *Suppose $a > 1$, $b \in (0,1)$, and $\frac{a^2}{d^2} + b^2\left(1 - \frac{1}{d^2}\right) \le 1$. Then $\mathcal{E}_{a,b} \subseteq B_X$.*

Proof. Let $b = \sqrt{\frac{d^2-a^2}{d^2-1}}$. Then $\psi(a) = ab^{n-1} = a\left(\frac{d^2-a^2}{d^2-1}\right)^{\frac{n-1}{2}}$, $\psi(1) = 1$. It is enough to show $\psi'(1) > 0$. Now

$$\psi'(a) = \left(\frac{d^2-a^2}{d^2-1}\right)^{\frac{n-1}{2}-1} \frac{d^2-na^2}{d^2-1},$$

which is indeed > 0 for $d > \sqrt{n}$. Note this is really a 2-dimensional argument; there is 1 special direction. We shrunk the y direction and expanded the x direction.

It's enough to show the new ellipse $\mathcal{E}_{a,b}$ is in the rhombus in the picture. Calculations are left to the reader.



□

□

2.3 Proof

Proof of Theorem 3.2.1. By translating f (so that it is 0 at some point), without loss of generality we can assume that for all $x \in \mathcal{N}_\delta$, $\|f(x)\|_Y \le 2L$.

Step 1: (Rough extension F_1 on S_X.) We show that there exists $F_1 : S_X \to Y$ such that for all $x \in \mathcal{N}_\delta \to Y$ such that for all $x \in \mathcal{N}_\delta$,

1. $\|F_1(x) - f(x)\|_Y \le 2L\delta$.
2. $\forall x, y \in S_X$, $\|F_1(x) - F_1(y)\| \le L(\|x-y\|_X + 4\delta)$.

This is a partition of unity argument. Write $\mathcal{N}_\delta = \{X_1, \ldots, X_N\}$. Consider

$$\{(x_p + 2\delta B_X) \cap S_X\}_{p=1}^N,$$

which is an open cover of S_X.

Let $\{\phi_p : S_X \to [0,1]\}_{p=1}^N$ be a partition of unity subordinted to this open cover. This means that

1. $\operatorname{Supp} \phi_p \subseteq x_p + 2\delta B_X$,
2. $\sum_{p=1}^N \phi_p(x) = 1$ for all $x \in S_X$.

For all $x \in S_X$, define $F_1(x) = \sum_{p=1}^N \phi_p(x) f(x_p) \in Y$. Then as F_1 is a weighted sum of $f(x_p)$'s, $\|F_1\|_\infty \le 2L$. If $x \in \mathcal{N}_\delta$, because $\phi_p(x)$ is 0 when $|x - x_p| > 2\delta$,

$$\|F_1(x) - f(x)\|_Y = \left\| \sum_{p: \|x-x_p\|_X \le 2\delta} \phi_p(x)(f(x_p) - f(x)) \right\| \le \sum_{\|x-x_p\| \le 2\delta} \phi_p(x) L \|x - x_p\|_X \le 2L\delta.$$

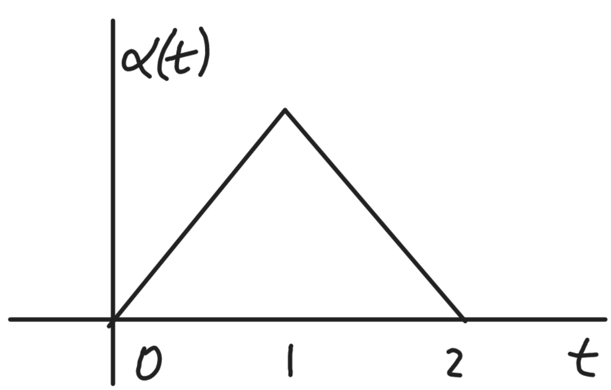
For $x, y \in S_X$,

$$\begin{aligned}
\|F_1(x) - F_1(y)\| &= \left\| \sum_{\substack{\|x-x_p\| \le 2\delta \\ \|y-x_q\| \le 2\delta}} (f(x_p) - f(x_q))\phi_p(x)\phi_q(x) \right\| \\
&\le \sum_{\substack{\|x-x_p\| \le 2\delta \\ \|y-x_q\| \le 2\delta}} L\|x_p - x_q\| \phi_p(x)\phi_p(y) \\
&\le L(\|x - y\| + 4\delta).
\end{aligned}$$

2-29: We continue proving Bourgain's almost extension theorem.
Step 2: Extend F_1 to F_2 on the whole space such that

1. $\forall x \in \mathcal{N}_\delta$, $\|F_2(x) - f(x)\|_Y \lesssim L\delta$.
2. $\|F_2(x) - F_2(y)\|_Y \lesssim L(\|x - y\|_X + \delta)$.
3. $\operatorname{Supp}(F_2) \subseteq 2B_X$.
4. F_2 is smooth.

Denote $\alpha(t) = \max\{1 - |1 - t|, 0\}$.



Let

$$F_2(x) = \alpha(\|x\|_X) F_1(x) \|x\|_X.$$

F_2 still satisfies condition 1. As for condition 2,

$$\begin{aligned}
\|F_2(x) - F_2(y)\|_Y &= \left\| \alpha(\|x\|_X) F_1\left(\frac{x}{\|x\|_X}\right) - \alpha(\|y\|_X) F_1\left(\frac{y}{\|y\|_X}\right) \right\| \\
&\le |\alpha(\|x\|) - \alpha(\|y\|)| \underbrace{\left\| F_1\left(\frac{x}{\|x\|_X}\right) \right\|}_{\le 2L} + \alpha(\|y\|) \left\| F_1\left(\frac{x}{\|x\|_X}\right) - F_1\left(\frac{y}{\|y\|_X}\right) \right\| \\
&\le (\|x\| - \|y\|) 2L + \alpha(\|y\|) L \left(\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| + 4\delta \right) \\
&\le 2L \|x - y\| + L\alpha(\|y\|) \left(\|x\| \left| \frac{1}{\|x\|} - \frac{1}{\|y\|} \right| + \frac{\|x-y\|}{\|y\|} + 4\delta \right) \\
&\le 2L \|x - y\| + L\alpha(\|y\|) \left(\frac{\|x-y\|}{\|y\|} + \frac{\|x-y\|}{\|y\|} + 4\delta \right) \\
&\lesssim L(\|x - y\| + \delta),
\end{aligned}$$

where in the last step we used $\alpha(\|y\|) \le \|y\|$ and $\alpha(\|y\|) \le 1$.

Note F_2 is smooth because the sum for F_1 was against a partition of unity and $\|\cdot\|_X$ is smooth, although we don't have uniform bounds on smoothness for F_2.

Step 3: We make F smoother by convolving.

Lemma 3.2.6 (Begun, 1999)**.** *Let $F_2 : X \to Y$ satisfy $\|F_2(x) - F_2(y)\|_Y \le L(\|x - y\|_X + \delta)$. Let $\tau \ge c\delta$. Define*

$$F(x) = \frac{1}{Vol(\tau B_X)} \int_{\tau B_X} F_2(x + y)\, dy.$$

Then

$$\|F\|_{Lip} \le L\left(1 + \frac{\delta n}{2\tau}\right).$$

The lemma proves the almost extension theorem as follows. We passed from $f : \mathcal{N}_\delta \to Y$ to F_1 to F_2 to F. If $x \in \mathcal{N}_\delta$,

$$\begin{aligned}
\|F(x) - f(x)\|_Y &= \left\| \frac{1}{\mathrm{Vol}(\tau B_X)} \int_{B_X} (F_2(x + y) - f(x))\, dy \right\| \\
&\le \frac{1}{\mathrm{Vol}(\tau B_X)} \int_{\tau B_X} \|F_2(x + y) - F_2(x)\|_Y + \underbrace{\|F_2(x) - f(x)\|_Y}_{\delta L}\, dy \\
&\le \frac{1}{\mathrm{Vol}(\tau B_X)} \int_{\tau B_X} (L(\underbrace{\|y\|_X}_{\le \tau} + \delta L))\, dy \lesssim L\tau.
\end{aligned}$$

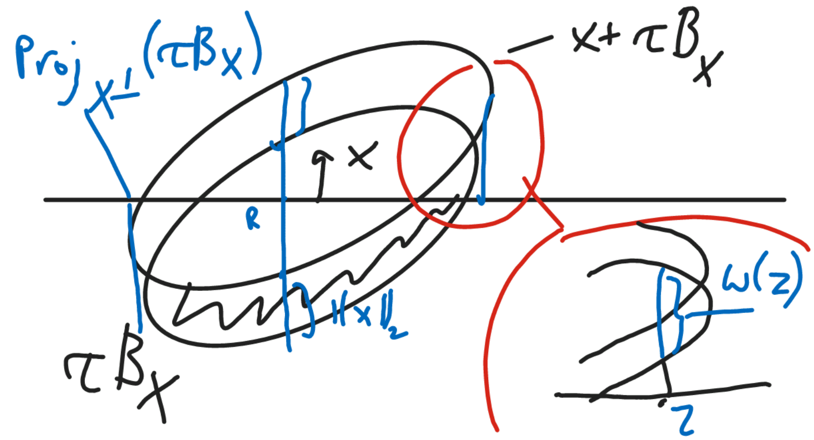
Now we prove the lemma.

Proof. We need to show

$$\|F(x)-F(y)\|_Y \le L\left(1+\frac{\delta n}{2\tau}\right)\|x-y\|_X .$$

Without loss of generality $y=0$, $\mathrm{Vol}(\tau B_X)=1$. Denote

$$M = \tau B_X \backslash (x+\tau B_X)$$
$$M' = (x+\tau B_X)\backslash \tau B_X .$$



We have

$$F(0)-F(x) = \int_M F_z(y)\,dy - \int_{M'} F_z(y)\,dy.$$

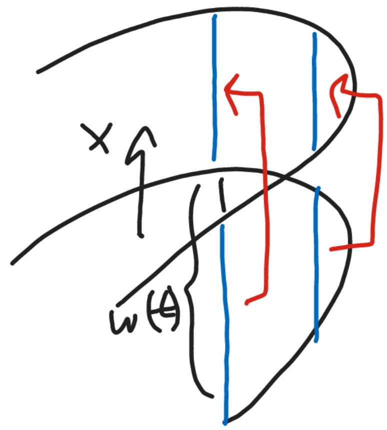
Define $\omega(z)$ to be the Euclidean length of the interval $(z+\mathbb{R}x)\cap(\tau B_X)$. By Fubini,

$$\int_{\mathrm{Proj}_{X^\perp}(\tau B_X)} \omega(z)\,dz = \mathrm{Vol}_n(\tau B_X) = 1.$$

Denote

$$W = \{z\in \tau B_X : (z+\mathbb{R}x)\cap(\tau B_X)\cap(x+\tau B_X)\neq\phi\}$$
$$N = \tau B_X\backslash W.$$

Define $C: M\to M'$ a shift in direction X on every fiber that maps the interval $(z+\mathbb{R}x)\cap M \to (z+\mathbb{R}x)\cap M'$.



C is a measure preserving transformation with

$$\|z - C(z)\|_X = \begin{cases} \|x\|_X, & z \le N \\ \omega(z)\frac{\|x\|_X}{\|x\|_2}, & z \in W \cap M. \end{cases}$$

(In the second case we translate by an extra factor $\frac{\omega(z)}{\|x\|_2}$.) Then

$$\begin{aligned}
\|F(0) - F(x)\|_Y &= \left\| \int_M F_2(y)\,dy - \int_{M'} F_2(y)\,dy \right\|_Y \\
&= \left\| \int_M (F_2(y) - F_2(C(y)))\,dy \right\|_Y \\
&\le \int_M L(\|y - C(y)\|_X + \delta)\,dy \\
&\le \int_M L(\|y - C(y)\|_X + \delta)\,dy \\
&= L\delta \mathrm{Vol}(M) + L\int_M \|y - C(y)\|_X \; dy \\
\int_M \|y - C(y)\|_X \; dy &= \int_N \|x\|_X \; dy + \int_{W\cap M} \frac{\omega(y)\,\|x\|_X}{\|x\|_2}\,dy \\
&= \|x\|_X \mathrm{Vol}(N) + \int_{\mathrm{Proj}(W\cap M)} \frac{\omega(z)\,\|x\|_X}{\|x\|_2}\,\|x\|_2\,dz \quad \text{orthogonal decomposition} \\
&= \|x\|_X \mathrm{Vol}(N) + \mathrm{Vol}(\tau B_X \backslash N) \\
&= \|x\|_X \mathrm{Vol}(\tau B_X) = \|x\|_X \,.
\end{aligned}$$

We show $M = \tau B_X \backslash (x + \tau B_X) \subseteq \tau B_X \backslash (1 - \frac{\|x\|}{\tau})\tau B_X$. Indeed, for $y \in M$,

$$\begin{aligned}
\|y - x\|_X &\ge \tau \\
\|y\| &\ge \tau - \|x\| = \left(1 - \frac{\|x\|}{\tau}\right)\tau.
\end{aligned}$$

□

Then

$$\mathrm{Vol}(M) \le \mathrm{Vol}(\tau B_X) - \mathrm{Vol}\left(\left(1 - \frac{\|x\|}{\tau}\right)\tau B_X\right) = 1 - \left(1 - \frac{\|x\|}{\tau}\right) \lesssim \frac{n\|x\|}{\tau}$$

□

Bourgain did it in a more complicated, analytic way avoiding geometry. Begun notices that careful geometry is sufficient.

Later we will show this theorem is sharp.

3 Proof of Bourgain's discretization theorem

At small distances there is no guarantee on the function f. Just taking derivatives is dangerous. It might be true that we can work with the initial function. But the only way Bourgain figured out how to prove the theorem was to make a 1-parameter family of functions.

3.1 The Poisson semigroup

Definition 3.3.1: The **Poisson kernel** is $P_t(x) : \mathbb{R}^n \to \mathbb{R}$ given by

$$P_t(x) = \frac{C_n t}{(t^2 + \|x\|_2^2)^{\frac{n+1}{2}}}, \quad C_n = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}}.$$

Proposition 3.3.2 (Properties of Poisson kernel)**:**

1. For all $t > 0$, $\int_{\mathbb{R}^n} P_t(x)\,dx = 1$.
2. (Semigroup property) $P_t * P_s = P_{t+s}$.
3. $\widehat{P_t}(x) = e^{-2\pi\|x\|_2 t}$.

Lemma 3.3.3. *Let F be the function obtained from Bourgain's almost extension theorem 3.2.1. For all $t > 0$, $\|P_t * F\|_{Lip} \lesssim 1$.*

Proof. We have

$$P_t * F(x) - P_t * F(y) = \int_{\mathbb{R}^n} P_t(z)(F(x-z) - F(x-y))\,dz.$$

Now use the fact that F is Lipschitz. □

Our goal is to show there exists $t_0 > 0$, $x \in B$ such that if we define

$$\text{eq:bdt-T}\; T = (P_{t_0} * F)'(x) : X \to Y, \tag{3.1}$$

(i.e., $Ta = \partial_a(P_{t_0} * F)(x)$). We showed $\|T\| \lesssim 1$. It remains to show $\|T^{-1}\| \lesssim D$. Then T has distortion at most $O(D)$, and hence T gives the desired extension in Bourgain's discretizaton theorem 1.2.4.

3-2: Today we continue with Bourgain's Theorem. Summarizing our progress so far:

1. Initially we had a function $f : \mathcal{N}_\delta \to Y$ and a δ-net $\mathcal{N}_\delta$ of B_X.
2. From f we got a new function $F : X \to Y$ with $\text{supp}(F) \subseteq 3B_X$, satisfying the following. (None of the constants matter.)
   (a) $\|F\|_{L_p} \le C$ for some constant.
   (b) For all $x \in \mathcal{N}_\delta$, $\|F(x) - f(x)\|_Y \le n\delta$.

Let $P_t : \mathbb{R}^n \to \mathbb{R}$ be the Poisson kernel. What can be said about the convolution $P_t * F$? We know that for all t and for all x, $\|(P_t * F)(x)\|_{X\to Y} \le 1$. The goal is to make this function invertible. More precisely, the norm of the inverted operator is not too small.

We'll ensure this one direction at a time by investigating the directional derivatives. For $g : \mathbb{R}^n \to Y$, the directional derivative is defined by: for all $a \in S_X$,

$$\partial_a g(x) = \lim_{t\to 0} \frac{g(x+ta) - g(x)}{t}.$$

There isn't any particular reason why we need to use the Poisson kernel; there are many other kernels which satisfy this. We *definitely* need the semigroup property, in addition to all sorts of decay conditions.

We need a lot of lemmas.

Lemma 3.3.4 (Key lemma 1)**.** lem:bdt-1 *There are universal constants C, C', C'' such that the following hold. Suppose $t \in (0, \frac{1}{2}]$, $R \in (0, \infty), \delta \in (0, \frac{1}{100n})$ satisfy*

$$\delta \le C\frac{t\log(3/t)}{\sqrt{n}} \le C'\frac{1}{n^{5/2}D^2} \tag{3.2}$$

$$C'' n^{3/2} D^2 \log(3/t) \le R \le \frac{C''}{t\sqrt{n}}. \tag{3.3}$$

Then for all $x \in \frac{1}{2}B_X, a \in S_X$, we have

$$(\|\partial_a(P_t * F)\|_Y * P_{Rt})(x) \ge \frac{C'''}{D}$$

This result says that given a direction a and look at how long the vector $\partial_a(P_t * F)$ is, it is not only large, but large on average. Here the average is over the time period Rt.

Note that I haven't been very careful with constants; there are some points where the constant is missing/needs to be adjusted.

The first condition says is that t is not too small, but the lower bound is extremely small. The upper bound is $\left(\frac{1}{n}\right)^k$ for some k whereas the lower bound is something like e^{-e^n}, so there's a huge gap.

The second condition says that $\log\left(\frac{1}{t}\right)$ is still exponential.

This lemma will come up later on and it will be clear.

From now on, let us assume this key lemma and I will show you how to finish. We will go back and prove it later.

Lemma 3.3.5 (Key lemma 2)**.** lem:bdt-2 *There's a constant C_1 such that the following holds (we can take $C_1 = 8$). Let μ be any Borel probability measure on S_X. For every $R, A \in (0, \infty)$, there exists $\frac{A}{(R+1)^{m+1}} \le t \le A$ such that*

$$\int_{S_X}\int_{\mathbb{R}^n} \|\partial_a(P_t * F)(x)\|_Y dx d\mu(a) \le \int_{S_X} \underbrace{\int_{\mathbb{R}^n} \|\partial_a(P_{(R+1)t} * F)\|_Y dx}_{(*)}\, d\mu(a) + \frac{C_1 \mathrm{vol}(3B_X)}{m}$$

Basically, under convolution, we can take the derivative under the integral sign. Thus (*) is an average of what we wrote on the left. Averaging only makes things smaller for norms, because norms are convex. Thus the right integral is less than the left integral, from Jensen's inequality.

The lemma says that adding that small factor, we get a bound in the opposite direction. We will argue that there must be a scale at which Jensen's inequality stabilizes, i.e., $\|\partial_a(P_t * F)\|_Y * P_{Rt}(x)$ stabilizes to a constant.

Note this is really the pigeonhole argument, geometrically.

Proof. Bourgain uses this technique a lot in his papers: find some point where the inequality is equality, and then leverage that.

If the result does not hold, then for every t in the range $\frac{t}{(R+1)^{m+1}} \le t \le A$, the reverse holds. We'll use the inequality at eveyr point in a geometric series. For all $k \in \{0, \dots, m+1\}$,

$$\int_{S_X}\int_{\mathbb{R}^n} \|\partial_a(P_{A(R+1)^{k-m-1}} * F)(x)\|_Y dx d\mu(a) > \int_{S_X}\int_{\mathbb{R}^n} \|\partial_a(P_{A(R+1)^{k-m}} * F)(x)\|_Y dx d\mu(a) + \frac{C_1 \mathrm{vol}(3B_X)}{m}.$$

Summing up these inequalities and telescoping

$$\int_{S_X}\int_{\mathbb{R}^n} \|\partial_a(P_{A(R+1)^{-m-1}} * F)(x)\|_Y dx d\mu(a) > \int_{S_X}\int_{\mathbb{R}^n} \|\partial_a(P_{A(R+1)} * F)(x)\|_Y dx d\mu(a) + \frac{8(m+1)\mathrm{vol}(3B_X)}{m} \tag{3.4}$$

(recall that we've been using the semigroup properties of the Poisson kernel this whole time). Now why is this conclusion absurd? Take C_1 to be the bound on the Lipschitz constant of F. Because $\|\partial_a F\|_Y \le 8$, we have$\partial_a(P_{A(R+1)^{-m-1}} * F)(x) \le C_1$. Since partial derivatives commute with the integral sign, we get

$$\int_{S_X}\int_{\mathbb{R}^n} \|\partial_a(P_{A(R+1)^{-m-1}} * F)(x)\|_Y \, dx \, d\mu(a) = \int_{S_X}\int_{\mathbb{R}^n} \|(P_{A(R+1)^{-m-1}} * \partial_a F)(x)\|_Y \, dx \, d\mu(a) \tag{3.5}$$

$$\le \int_{S_X}\int_{\mathbb{R}^n} C_1 \le C_1 \mathrm{vol}(B_X) \tag{3.6}$$

because F is Lipschitz and the Poisson semigroup integrates to unity. Together (3.4) and (3.6) give a contradiction. □

Now assuming Lemma 3.3.4 (Key Lemma 1), let's complete the proof of Bourgain's discretization theorem. Assume from now on $\delta < \left(\frac{1}{cD}\right)^{(CD)^{2n}}$ where C is a large enough constant that we will choose ($C = 500$ or something).

Proof of Bourgain's Discretization Theorem 1.2.4. Let $\mathcal{F} \subseteq S_X$ be a $\frac{1}{C_2 D}$-net in S_X. Then $|\mathcal{F}| \le (C_3 D)^n$ for some C_3. We will apply the Lemma 3.3.5 (Key Lemma 2) with μ the uniform measure on $\mathcal{F}$,

$$\begin{aligned} A &= (1/CD)^{5n} \\ R+1 &= (CD)^{4n} \\ m &= \lceil (CD)^{n+1} \rceil. \end{aligned}$$

Then there exists $(1/(CD))^{(CD)^{2n}} \le t \le (1/(CD))^{5n}$ such that

$$\text{eq:bdt-pf-lem2} \sum_{a\in\mathcal{F}} \int_{\mathbb{R}^n} \|\partial_a (P_t * F)(x)\|_Y dx \le \sum_{a\in\mathcal{F}} \int_{\mathbb{R}^n} \|\partial_a (P_{(R+1)t} * F)(x)\|_Y dx + \frac{8\mathrm{vol}(3B_X)}{m} |\mathcal{F}| \tag{3.7}$$

We check the conditions of Lemma 3.3.4 (Key Lemma 1).

1. For (3.2), note t is exponentially small, so the RHS inequality is satisfied. For the LHS inequality, note $\delta \le t$ and $\frac{C\ln\left(\frac{1}{t}\right)}{\sqrt{n}} \le 1$. To see the second inequality, note $\frac{C\ln\left(\frac{1}{t}\right)}{\sqrt{n}} \ge \frac{C5n\ln(CD)}{\sqrt{n}} \ge 1$ (for n large enough).

2. For (3.2), note the LHS is dominated by $\ln\left(\frac{1}{t}\right) \le (CD)^{2n}\ln(CD)$ which is much less than $R = (CD)^{4n} - 1$, and $\frac{1}{t}$, the dominating term on the RHS, is $\ge (CD)^{5n}$.

In my paper (reference?), I wrote what the exact constants are. They're written in the paper, and are not that important. We're choosing our constants big enough so that our inequalities hold true with room to spare.

Now we can use Key Lemma 1, which says

$$\text{eq:bdt-pf-lem1} \left(\|\partial_a (P_t * F)\|_Y * P_{Rt}\right)(x) \ge \frac{1}{D} \tag{3.8}$$

Using $P_{(R+1)t} = P_{Rt} * P_t$ (semigroup property) and Jensen's inequality on the norm, which is a convex function, we have

$$\text{eq:bdt-conv} \|\partial_a (P_{(R+1)t} * F)(x)\|_Y = \|\left(\partial_a (P_t * F)\right) * P_{Rt}(x)\|_Y \le \left(\|\partial_a (P_t * F)\|_Y * P_{Rt}\right)(x). \tag{3.9}$$

Since the norm is a convex function, by Jensen's we get our inequality.

Let

$$\psi(x) := \left(\|\partial_a (P_t * F)\|_Y * P_{Rt}\right)(x) - \|\partial_a \left(P_{(R+1)t} * F\right)(x)\|_Y$$

From (3.9), $\psi(x) \ge 0$ pointwise. Using Markov's inequality,

$$\mathrm{vol}\left\{x \in \frac{1}{2}B_X : \psi(x) > \frac{1}{D}\right\} \le D \int_{\mathbb{R}^n} \left(\left(\|\partial_a (P_t * F)\|_Y * P_{Rt}\right)(x) - \|\partial_a \left(P_{(R+1)t} * F\right)(x)\|_Y\right) dx.$$

because we can upper bound the integral over the ball by an integral over $\mathbb{R}^n$. Note we are using the probabilistic method.

This inequality was for a fixed a. We now use the union bound over $\mathcal{F}$, a δ-net of a's to get

$$\text{vol}\left\{x \in \frac{1}{2}B_X : \exists a \in \mathcal{F}, \psi_a(x) > 1/D\right\} \tag{3.10}$$
$$\le \sum_{a\in\mathcal{F}} \text{vol}(x \in \frac{1}{2}B_X : \psi_a(x) > 1/D) \tag{3.11}$$
$$\le D \sum_{a\in\mathcal{F}} \int_{\mathbb{R}^n} \left(\|\partial_a(P_t * F)\|_Y * P_{Rt})(x) - \|\partial_a(P_{(R+1)t} * F)(x)\|_Y\right) dx \tag{3.12}$$
$$= D \sum_{a\in\mathcal{F}} \int_{\mathbb{R}^n} \left(\|\partial_a(P_t * F)\|_Y)(x) - \|\partial_a(P_{(R+1)t} * F)(x)\|_Y\right) dx \tag{3.13}$$
$$\le \frac{C_1 \text{vol}(3B_X)}{m}(C_3 D)^n < \text{vol}\left(\frac{1}{2}B_X\right). \tag{3.14}$$

where (3.13) follows because when you convolve something with a probability measure, the integral over $\mathbb{R}^n$ does not change, and (3.14) follows from (3.7) and our choice of m.

We've proved that there must exist a point in half the ball such that for every $a \in \mathcal{F}$, our net, we have

$$\frac{1}{D} \ge \|\partial_a(P_t * F)\|_Y * P_{Rt}(x) - \|\partial_a(P_{(R+1)t} * F)(x)\|_Y$$

I think we want either this to be $\frac{1}{2D}$, or (3.8) to be $\frac{1}{D/2}$, in order to get the following bound (with $\frac{1}{2D}$ or $\frac{1}{D}$). This involves changing some constants in the proof.

From this we can conclude that for this x,

$$\|Ta\|_Y = \|\partial_a(P_{(R+1)t} * F)(x)\|_Y \ge 1/D$$

where $a \in S_X$ we let $t_0 = (R+1)t$. This shows $\|T^{-1}\| \le D$. □

For the exact constants, see the paper (reference).

Note this is very much a probabiblistic existence statement or result. Usually we estimate by hand the random variable we want to care about. Here we want to prove an existential bound, so we estimate the probability of the bad case. But we estimate the probability of the bad case using another proof by contradiction.

It remains to estimate some integrals.

3-7: Finishing Bourgain. There's a remaining lemma about the Poisson semigroup that we're going to do today (Lemma 3.3.4); it's nice, but it's not as nice as what came before.

The last remaining lemma to prove is Lemma 3.3.4.

Remember that $\frac{1}{\sqrt{n}}\|x\|_2 \le \|x\|_X \le \|x\|_2$ for all $x \in X$. We need three facts about the Poisson kernel.

Proposition 3.3.6 (Fact 1): pr:pois1

$$\int_{\mathbb{R}^n \setminus (rB_X)} P_t(x) dx \le \frac{t\sqrt{n}}{r}.$$

Proof. First note that the Poisson semigroup has the property that

$$\text{eq:pois-rescale} P_t(x) = \frac{1}{t^n} P_1(x/t), \tag{3.15}$$

i.e., P_t is just a rescaling of P_1. So it's enough to prove this for $t = 1$. We have

$$\begin{aligned}
\int_{\|x\|_X \ge r} P_t(x)\,dx &\le \int_{\|x\|_2 \ge r} P_t(x)\,dx \\
&= \int_{\|x\|_2 \ge r/t} P_1(x)\,dx \\
&\quad \text{by change of variables and (3.15)} \\
&= C_n S_{n-1} \int_{r/t}^{\infty} \frac{s^{n-1}}{(1+s^2)^{(n+1)/2}} ds \\
&\quad \text{polar coordinates, where } S_{n-1} = \text{vol}(\mathbb{S}^{n-1}) \\
&\le C_n S_{n-1} \int_{r/t}^{\infty} \frac{1}{s^2} ds \\
&= \frac{t}{r} C_n S_{n-1} \\
&\le \frac{t\sqrt{n}}{r},
\end{aligned}$$

where the last inequality follows from expanding in terms of the Gamma function and using Stirling's formula.

Above we changed to polar coordinates using

$$\int_{\|x\|_2 \le R} f(\|x\|)\,dx = \int_0^R S_{n-1} \|x\|^{n-1} f(r)\,dr.$$

□

Proposition 3.3.7 (Fact 2)**:** For all $y \in \mathbb{R}^n$,

$$\int_{\mathbb{R}^n} |P_t(x) - P_t(x+y)| dx \le \frac{\sqrt{n}\|y\|_2}{t}.$$

Proof. It's again enough to prove this when $t = 1$.

$$\begin{aligned}
\int_{\mathbb{R}^n} |P_t(x) - P_t(x+y)|\,dx &= \int_{\mathbb{R}^n} \left| \int_0^1 \langle \nabla P_t(x+sy), y \rangle ds \right| dx \\
&\le \|y\|_2 \int_{\mathbb{R}^n} \|\nabla P_t(x)\|_2 dx && \text{Cauchy-Schwarz} \\
&\le \|y\|_2 (n+1) C_n \int_{\mathbb{R}^n} \frac{\|x\|_2}{(1+\|x\|_2^2)^{\frac{n+3}{2}}} dx && \text{computing gradient} \\
&= \|y\|_2 (n+1) C_n S_{n-1} \int_0^{\infty} \frac{r^n}{(1+r^2)^{\frac{n+3}{2}}}\,dr && \text{polar coordinates} \\
&\le \sqrt{n}\, \|y\|_2 .
\end{aligned}$$

The integral and multiplicative constants perfectly cancels out the $n+1$ and becomes 1 (calculation omitted). $\square$

Proposition 3.3.8 (Fact 3)**:** For all $0 < t < \frac{1}{2}$ and $x \in B_X$, we have

$$\|P_t * F(x) - F(x)\|_Y \le \sqrt{n}t \log(3/t).$$

Proof. The LHS equals

$$\left\|\int_{\mathbb{R}^n} (F(x-y) - F(x))\, P_t(y)dy\right\|_Y \le \underbrace{\int_{x+3B_X} \|F(x-y) - F(x)\|_Y\, P_t(y)dy}_{(3.16)} + c\underbrace{\int_{\mathbb{R}^n \setminus (x+3B_X)} P_t(y)dy}_{(3.17)}.$$

Using $\|F(x-y) - F(x)\|_Y \le \|F\|_{\text{Lip}} \frac{c}{\rho} \cdot \|y\|_X$ and that $\|F\|_{\text{Lip}}$ is a constant,

$$\begin{aligned}
(3.16) &= \int_{x+3B_X} \|y\|_X P_t(y)dy \\
&\le \int_{4\sqrt{n}B_{l_2^n}} \|y\|_2 P_t(y)dy && \text{using } x \in B_X \implies x + 3B_X \subseteq 4\sqrt{n}B_{l_2} \\
&= tC_n S_{n-1} \int_0^{\frac{4\sqrt{n}}{t}} \frac{s^n}{(1+s^2)^{(n+1)/2}}\, ds && \text{polar coordinates} \\
&\le t\sqrt{n}\left(\int_0^{\sqrt{n}} \frac{1}{\sqrt{n}}\, ds + \int_{\sqrt{n}}^{4\sqrt{n}/t} \frac{1}{s} ds\right) \\
&\le t\sqrt{n}\left(1 + \ln\left(\frac{4}{t}\right)\right),
\end{aligned}$$

where we used the fact that $\frac{s^n}{(1+s^2)^{\frac{n+1}{2}}}$ is maximized when $s = \sqrt{n}$, and is always $\le \frac{1}{s}$.

For the second term, note that $2B_X \subseteq x + 3B_X$, so

$$(3.17) = c\int_{\mathbb{R}^n \setminus 2B_X} P_t(y) \le \frac{ct\sqrt{n}}{2}$$

where we used the tail bound of Fact 1 (Pr. 3.3.7). Adding these two bounds gives the result. $\square$

Now we have these three facts, we can finish the proof of the lemma.

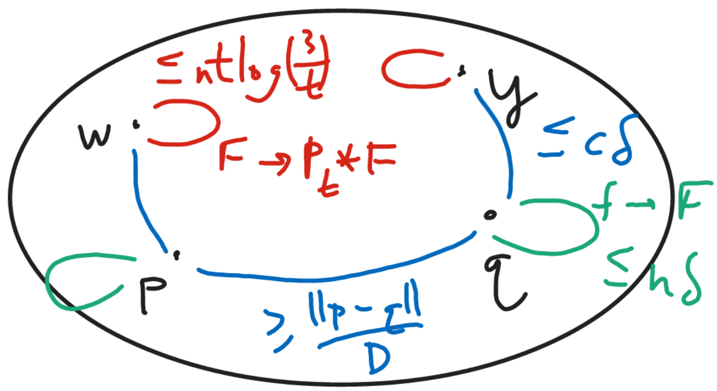
Proof of Lemma 3.3.4.

Claim 3.3.9 ($P_t * F$ separates points)**.** *Let* $\theta = cD\sqrt{n}t\log(3/t)$. *Let* $w, y \in \frac{1}{2}B_X$ *with* $\|w - y\|_X \ge \theta$. *Then*

$$\|P_t * F(w) - P_t * F(y)\|_Y \ge 1/D$$

We do not claim at all that these numbers are sharp.

Proof. We can find a point $p \in \mathcal{N}_\delta$ which is δ away from w, and $q \in \mathcal{N}_\delta$ which is δ away from y. We also know that F is close to f, and we can use Fact 3 (Pr. 3.3.8) to bound the error of convolving.



By the triangle inequality

$$\begin{aligned}
\|P_t * F(w) - P_t * F(y)\|_Y &\geq \|f(p) - f(q)\| - \|F(p) - f(p)\| - \|F(q) - f(q)\| \\
&\quad - \|F(w) - F(p)\| - \|F(y) - F(q)\| \\
&\quad - \|P_t * F(w) - F(w)\| - \|P_t * F(y) - F(y)\| \\
&\geq \frac{\|p-q\|}{D} - 2n\delta - c\delta - \sqrt{n}t\log(3/t) \\
&\geq \frac{\|y-w\| - 2\delta}{D} - 2n\delta - c\delta - \sqrt{n}t\log(3/t).
\end{aligned}$$

Taking

$$\theta = cD\sqrt{n}t\log(3/t),$$

we find this is at least a constant times $\|y-w\|/D$, we need $\theta = cD\sqrt{n}t\log(3/t)$. Note that $\sqrt{n}t\log(3/t)$ is the largest error term because by the assumptions in the lemma, it is greater than $2n\delta$. □

Now consider

$$\|P_t * F(z + \theta a) - P_t * F(z)\|$$

By the claim, if $z \in \frac{1}{4}B_X$, then we know

$$\begin{aligned}
\theta/D &\leq \|P_t * F(z+\theta a) - P_t * F(z)\| && (3.18)\\
&= \int_0^\theta \|\partial_a(P_t * F)(z + sa)\|_Y && (3.19)
\end{aligned}$$

Suppose $\|x-y\| \le \frac{1}{4}$ (we will apply (3.19) to $x-y=z$), $x \in \frac{1}{2}B_X$, $y \in \frac{1}{8}B_X$. By throwing away part of the integral,

$$\frac{1}{\theta}\int_0^\theta \int_{\mathbb{R}^n} \|\partial_a (P_t * F)(x+sa-y)\|_Y P_{Rt}(y)dy \tag{3.20}$$

$$\ge \frac{1}{\theta}\int_0^\theta \int_{(1/8)B_X} \|\partial_a (P_t * F)(x-y+sa)\|_Y P_{Rt}(y)\,ds\,dy \tag{3.21}$$

$$= \int_{(1/8)B_X} \left(\frac{1}{\theta}\int_0^\theta \|\partial_a (P_t * F)(x-y+sa)\|_Y\,ds\right) P_{Rt}(y)\,dy \qquad \text{by Fubini} \tag{3.22}$$

$$\ge \frac{1}{D}\int_{(1/8)B_X} P_{Rt}(y)dy \qquad \text{by (3.19)} \tag{3.23}$$

$$\ge \frac{c}{D} \tag{3.24}$$

for some constant c. The calculation above says that the average length in every direction is big, and if you average the average length again, it is still big. How do we get rid of the additional averaging? We care about

$$\begin{aligned}
\|\partial_a(P_t*F)\| * P_{Rt}(x) &= \int_{\mathbb{R}^n} \|\partial_a(P_t*F)(x-y)\|_Y P_{Rt}(y)dy \\
&= \frac{1}{\theta}\int_0^\theta \left(\int_{\mathbb{R}^n} \|\partial_a(P_t*F)(x-y+sa)\|_Y P_{Rt}(y)dy\right) \\
&\quad - \frac{1}{\theta}\int_0^\theta \int_{\mathbb{R}^n} \|\partial_a(P_t*F)(x-y)\|_Y (P_{Rt}(y+sa)-P_{Rt}(y))dy \\
&> \frac{c}{D} - \frac{c'}{\theta}\int_0^\theta \int_{\mathbb{R}^n} |P_{Rt}(y+sa)-P_{Rt}(y)|ds\,dy && \text{by (3.24) and what?} \\
&\ge \frac{c}{D} - \frac{c'}{\theta}\int_0^\theta \frac{\sqrt{n}s\|a\|_2}{Rt}ds && \text{by Fact 2 (Pr. 3.3.7)}
\end{aligned}$$

This error term is

$$\frac{c'}{\theta}\int_0^\theta \frac{\sqrt{n}s\|a\|_2}{Rt}ds = \frac{cDn\log\left(\frac{3}{t}\right)}{R} = O\left(\frac{1}{D}\right)$$

by (3.2), so $\|\partial_a(P_t*F)\|*P_{Rt}(x) = \Omega\left(\frac{1}{D}\right)$ (if we chose constants correctly), as desired. Factor of $\sqrt{n}$? □

To summarize: You compute the error such that you have distance θ. Then you look at the average derivative on each line like this. Then, the average derivative is roughly $\|\partial_a(P_t*F)\|*P_{Rt}(x)$. So averaging the derivative in a bigger ball is the same up to constants as taking a random starting point and averaging over the ball. What's written in the lemma is that the derivative of a given point in this direction, averaged over a little bit bigger ball, it's not unreasonable to see that at first take starting point and going in direction θ and averaging in the bigger ball is equivalent to randomizing over starting points.

3-21: A better bound in Bourgain Discretization

4 Upper bound on $\delta_{X\to Y}$

We give an example where we need a discretization bound δ going to 0.

Theorem 3.4.1. *There exist Banach spaces X, Y with dim $X = n$ such that if $\delta < 1$ and $\mathcal{N}_\delta$ is a δ-net in the unit ball and $C_Y(\mathcal{N}_\delta) \gtrsim C_Y(X)$, then $\delta \lesssim \frac{1}{n}$.*

Together with the lower bound we proved, this gives

$$\frac{1}{\rho n^{Cn}} < \delta_{X\to Y}(1/2) < \frac{1}{n}$$

The lower bound would be most interesting to improve.

The example will be $X = \ell_1^n, Y = \ell_2$.

We need two ingredients.

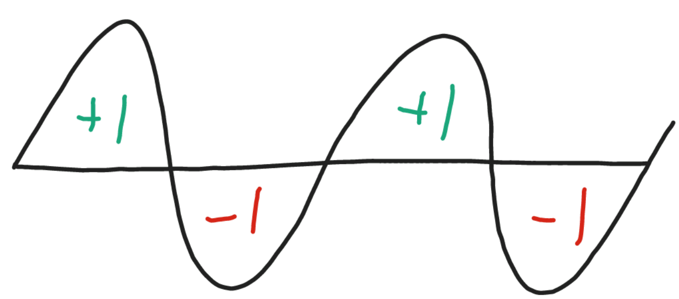
Firstly, we show that $(L_1, \sqrt{\|x-y\|_1})$ is isometric to a subset of L_2. More generally, we prove the following.

Lemma 3.4.2. *Given a measure space (Ω, μ), $(L_1(\mu), \sqrt{\|x-y\|_1})$ is isometric to a subset of $L_2(\Omega \times \mathbb{R}, \mu \times \lambda)$, where λ is the Lesbegue measure of $\mathbb{R}$.*

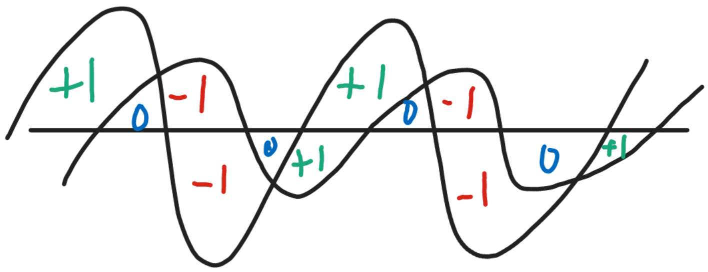
Note all separable Hilbert spaces are the same, $L_2(\Omega \times \mathbb{R}, \mu \times \lambda)$ is the same whenever it is separable.

Proof. First we define $T : L_1(\mu) \to L_2(\mu \times \lambda)$ as follows. For $f \in L_1(\mu)$, let

$$Tf(w,x) = \begin{cases} 1 & 0 \le f(w) \le x \\ -1 & x \le f(x) \le 0 \\ 0 & \text{otherwise} \end{cases}$$



Consider two functions $f_1, f_2 \in L_1(\mu)$. The function $|Tf_1 - Tf_2|$ is the indicator of the area between the graph of f_1 and the graph of f_2.



Thus the L_2 norm is, letting $\mathbb{1}$ be the signed indicator,

$$\begin{aligned}\|Tf_1 - Tf_2\|_{L_2(\mu\times\lambda)} &= \sqrt{\int_\Omega \int_{\mathbb{R}} \left(\mathbb{1}_{f_1(w),f_2(w)}(x)^2\right)\,dx\,d\mu} \\ &= \sqrt{\int_\Omega |f_1(w) - f_2(w)|\,d\mu(w)}\end{aligned}$$

□

More generally, if $q \le p < \infty$, then $(L_q, \|x-y\|_q^{p/q})$ is isometric to a subset of L_p. We may prove this later if we need it.

The second ingredient is due to Enflo .

Theorem 3.4.3 (Enflo, 1969)**.** *The best possible embedding of the hypercube into Hilbert space has distortion* $\sqrt{n}$*:*

$$C_2(\{0,1\}^n, \|\cdot\|_1) = \sqrt{n}$$

Note that we don't just calculate C_2 up to a constants here; we know it exactly. This is very rare.

Proof. We need to show an upper bound and a lower bound.

1. To show the upper bound, we show the identity mapping $\{0,1\}^n \to \ell_2^n$ has distortion $\sqrt{n}$. We have

$$\|x-y\|_2 = \left(\sum_{i=1}^n |x_i - y_i|^2\right)^{1/2} = \left(\sum_{i=1}^n |x_i - y_i|\right)^{1/2} = \sqrt{\|x-y\|_1},$$

since the values of x, y are only $0, 1$. Then

$$\frac{1}{\sqrt{n}}\|x-y\|_1 \le \|x-y\|_2 = \sqrt{\|x-y\|_1} \le \|x-y\|_1$$

Thus $C_2(\{0,1\}^n) \le \sqrt{n}$.

(So this theorem tells us that if you want to represent the boolean hypercube as a Euclidean geometry, nothing does better than the identity mapping.)

2. For the lower bound $C_2(\{0,1\}^n, \|\cdot\|_1) \geq \sqrt{n}$, we use the following.

   Lemma 3.4.4 (Enflo's inequality)**.** *We claim that for every $f : \{0,1\}^n \to \ell_2$,*

   $$\sum_{x \in \mathbb{F}_2^n} \|f(x) - f(x+e)\|_2^2 \leq \sum_{j=1}^{n} \sum_{x \in \mathbb{F}_2^n} \|f(x+e_j) - f(x)\|_2^2,$$

   where $e = (1, \ldots, 1)$ and e_i are standard basis vectors of $\mathbb{R}^n$.

   Here, addition is over $\mathbb{F}_2^n$.

   This is specific to ℓ_2. See the next subsection for the proof.

   In other words, Hilbert space is of Enflo-type 2 (Definition 1.1.13): the sum of the squares of the lengths of the diagonals is at most the sum of the squares of the lengths of the edges.

   Suppose
   $$\frac{1}{D}\|x-y\|_1 \leq \|f(x) - f(y)\|_2 \leq \|x-y\|_1.$$
   Our goal is to show that $D \geq \sqrt{n}$. Letting $y = x + e$,

   $$\|f(x+e) - f(x)\|_2 \geq \frac{1}{D}\|(x+e) - x\|_1 = \frac{n}{D}.$$

   Plugging into Enflo's inequality 3.4.4, we have, since there are 2^n points in the space,

   $$2^n(n/D)^2 \leq n \cdot 2^n \cdot 1^2 \implies D^2 \geq n$$

   as desired.

□

Proof of Theorem 3.4.1. Let $\mathcal{N}_\delta$ be a δ-net in $B_{\ell_1^n}$. Let $T : \ell_1^n \to L_2$ satisfy $\|T(x) - T(y)\|_2 = \sqrt{\|x-y\|_1}$. T restricted to $\mathcal{N}_\delta$ has distortion $\leq \sqrt{\frac{2}{\delta}}$ because for $x, y \in \mathcal{N}_\delta$ with $x \neq y$, we have
$$\frac{1}{\sqrt{2}}\|x-y\|_1 \leq \sqrt{\|x-y\|_1} \leq \frac{1}{\sqrt{\delta}}\|x-y\|_1.$$
However, the distortion of ℓ_1^n in ℓ_2 is $\sqrt{n}$. The condition $C_Y(\mathcal{N}_\delta) \gtrsim C_Y(X)$ means that for some constant K,
$$\sqrt{\frac{2}{\delta}} \geq K\sqrt{n} \implies \delta \leq \frac{2}{K^2 n}.$$

□

4.1 Fourier analysis on the Boolean cube and Enflo's inequality

We can prove this by induction on n (exercise). Instead, I will show a proof that is Fourier analytic and which generalizes to other situations.

Definition 3.4.5 (Walsh function)**:** Let $A \subseteq \{1, \ldots, n\}$. Define the Walsh function

$$W_A : \mathbb{F}_2^n \to \{\pm 1\}$$

by

$$W_A(x) = (-1)^{j \in A}.$$

Proposition 3.4.6 (Orthonormality)**:** For $A, B \subseteq \{1, \ldots, n\}$.

$$\mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n} W_A(x) W_B(x) = \delta_{AB}.$$

Thus, $\{W_A\}_{A \subseteq \{1,\ldots,n\}}$ is an orthonormal basis of $L_2(\mathbb{F}_2^n)$.

Proof. Without loss of generality $j \in A \backslash B$. Then

$$\mathbb{E}\,(-1)^{\sum_{j\in A} x_j + \sum_{j \in B} x_j} = 0.$$

□

Corollary 3.4.7. *Let $f : \mathbb{F}_2^n \to X$ where X is a Banach space. Then*

$$f = \sum_{A \subseteq \{1,\cdots,n\}} \hat{f}(a) W_A$$

where $\hat{f}(A) = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} f(x)(-1)^{\sum_{j \in A} x_j}$.

Proof. How do we prove two vectors are the same in a Banach space? It is enough to show that any composition with a linear functional gives the same value; thus it suffices to prove the claim for $X = \mathbb{R}$. The case $X = \mathbb{R}$ holds by orthonormality. □

Proof of Lemma 3.4.4. It is sufficient to prove the claim for $X = \mathbb{R}$; we get the inequality for ℓ_2 by summing coordinates. (Note this is a luxury specific to ℓ_2.)

The summand of the LHS of the inequality is

$$\begin{aligned} f(x) - f(x+e) &= \sum_{A \subseteq \{1,\ldots,n\}} \hat{f}(A)\left(W_A(x) - W_A(x+e)\right) \\ &= \sum_{A \subseteq \{1,\ldots,n\}} \hat{f}(A) W_A(x) \left(1 - (-1)^{|A|}\right) \\ &= \sum_{A \subseteq \{1,\ldots,n\}, |A| \text{ odd}} 2\hat{f}(A) W_A(x) \end{aligned}$$

The summand of the RHS of the inequality is

$$\begin{aligned}
-f(x+e_j)+f(x) &= \sum_{A\subseteq\{1,\ldots,n\}} \hat{f}(A)(W_A(x)-W_A(x+e_j)) \\
&= \sum_{A\subseteq\{1,\ldots,n\},j\in A} 2\hat{f}(A)W_A(x).
\end{aligned}$$

Summing gives

$$\begin{aligned}
\sum_{x\in\mathbb{F}_2^n}(f(x)-f(x+e))^2 &= 2^n\left\|\sum_{|A|\text{ odd}} 2\hat{f}(A)W_A\right\|_{L_2(\mathbb{F}_2^n)}^2 \\
&= 2^n\sum_{|A|\text{ odd}} 4(\hat{f}(A))^2 \\
\sum_{j=1}^n\sum_{x\in\mathbb{F}_2^n}(f(x+e_j)-f(x))^2 &= \sum_{j=1}^n 2^n\sum_{A:j\in A} 4\left(\hat{f}(A)\right)^2 \\
&= 2^n\sum_A\sum_{j\in A} 4\hat{f}(A)^2 \\
&= 2^n\sum_A 4|A|\hat{f}(A)^2
\end{aligned}$$

From this we see that the claim is equivalent to

$$\sum_{A\subseteq\{1,\ldots,n\},|A|\text{ odd}} \hat{f}(A)^2 \le \sum_{A\subseteq\{1,\ldots,n\}} |A|\hat{f}(A)^2$$

which is trivially true. □

This inequality invites improvements, as this is a large gap. The Fourier analytic proof made this gap clear.

5 Improvement for L_p

Theorem 3.5.1. thm:bourgain-lp *Suppose $p \ge 1$ and X is an n-dimensional normed space. Then*

$$\delta_{X\hookrightarrow L_p}(\varepsilon) \gtrsim \frac{\epsilon^2}{n^{5/2}}$$

In the world of embeddings into L_p, we are in the right ballpark for the value of δ—we know it is a power of n, if not exactly $1/n$. (It is an open question what the right power is.) This is a special case of a more general theorem, which I won't state. This proof doesn't use many properties of L_p.

Proof. This follows from Theorem 3.5.2 and Theorem 3.5.3. □

We will prove the theorem for $\varepsilon = \frac{1}{2}$. The proof needs to be modified for the ε-version.

Theorem 3.5.2. thm:bourgain-lp2 *Let X, Y be Banach spaces, with $\dim X = n < \infty$. Suppose that there exists a universal constant c such that $\delta \le \frac{c\varepsilon^2}{n^{5/2}}$, and that $\mathcal{N}_\delta$ is a δ-net of B_X.* [3] *Then there exists a separable probability space (Ω, μ)*[4]*, a finite dimensional subspace $Z \subseteq Y$ and a linear operator $T : X \to L_\infty(\mu, Z)$ such that for every $x \in X$ with $\|x\|_X = 1$,*

$$\frac{1-\varepsilon}{D} \le \int_\Omega \|Tx(\omega)\|_Y d\mu(\omega) \le \text{esssup}_{\omega\in\Omega}\|(Tx)\omega\|_Y \le 1+\epsilon$$

Note that the inequality can also be written as

$$\frac{1-\varepsilon}{D}\|Tx\|_{L_1(\mu,Z)} \le \|Tx\|_{L_\infty(\mu,Z)}.$$

Now what happens when $Y = L_p([0,1])$? We have

$$L_p(\mu, Z) \subseteq L_p(\mu, Y) = L_p(\mu, L_p) = L_p(\mu \times \lambda)$$

A different way to think of this is that a function $f : X \to L_\infty(\mu, L_p([0,1]))$ can be thought of as a function of $x \in X$ and $\omega \in [0,1]$, $f(\omega, x) : \Omega \times [0,1] \to \mathbb{R}$.

This is a very classical fact in measure theory:

Theorem 3.5.3 (Kolmogorov's representation theorem)**.** *Any separable $L_p(\mu)$ space is isometric to a subset of L_p.*

This is an immediate corollary of the fact that if I give you a separable probability space, there is a separable transformation of the probability space $(\Omega, \mu) \cong ([0,1], \lambda)$ where λ is Lebesgue measure. So there is a measure preserving isomorphism if the probability measure is atom-free. In general, the possible cases are

- $(\Omega, \mu) \cong ([0,1], \lambda)$ if it is atom free.
- $(\Omega, \mu) \cong [0,1] \times \{1, \dots, n\}$ if there are finitely many (n) atoms
- $(\Omega, \mu) \cong [0,1] \times \mathbb{N}$ if there are countably many atoms,
- $(\Omega, \mu) \cong \{1, \dots, n\}$ or $\mathbb{N}$ if it's completely atomic.

See Halmos's book for deails. This is just called Kolmogorov's representation theorem.

Now for this to be an improvement of Bourgain's discretization theorem for L_p, we need that if $Z \subset Y$ is a finite dimensional subspace and $W \subset L_\infty(\mu, Z)$ is a finite dimensional subspace such that on W, the L_∞ and L_1 norms are equivalent (a very strong restriction), then W embeds in Y. That's the only property of L_p we're using.

However, not every Y has this property.

[3] Let me briefly say something about vector-valued L_p spaces. Whenever you write $L_p(\mu, Z)$, this equals all functions $f : \Omega \to Z$ such that $\left(\int \|f(\omega)\|_Z^p d\mu(\omega)\right)^{1/p} < \infty$ (our norm is bounded).

[4] it will in the end be a uniform measure on half the ball

Theorem 3.5.4 (Ostrovskii and Randrianantoanina)**.** *Not every Y has this property.*

You can construct spaces of functions taking values in a finite dimensional subspace of it which cannot be embedded back into Y. Nevertheless, you have to work to find bad counterexamples. This is not published yet.

Next class we'll prove Theorem 3.5.2, which implies our first theorem. What is the advantage of formulating the theorem in this way? We can use Bourgain's almost-extention theorem along the ball. The function is already Lipschitz, so we can already differentitate it. The measure itself will be the unit ball. We'll look at the derivative of the extended function in direction ω at point x. That is what our T will be. We will look in all possible directions. Then the lower bound we have here says that the derivative of the function is only invertible on average, not at every point in the sphere. That tends to be big on average. We will work for that, but you see the advantage: In this formulation, we're going to get an average lower-bound, not an always lower-bound, and in this way we will be able to shave off two exponents. But this is not for Banach spaces in general. The advantage we have here is that L_p of L_p is L_p.

3/23: We will prove the improvement of Bourgain's Discretization Theorem.

Proof of Theorem 3.5.2. There exists $f : \mathcal{N}_\delta \to Y$ such that

$$\frac{1}{D}\|x-y\|_Y \le \|f(x)-f(y)\|_Y \le \|x-y\|_X$$

for $x, y \in \mathcal{N}_\delta$. By Bourgain's almost extension theorem, letting $Z = \operatorname{span}(f(\mathcal{N}_\delta))$, there exists $F : X \to Z$ such that

1. $\|F\|_{L_p} \lesssim 1$.
2. For every $x \in \mathcal{N}_\delta$, $\|F(x) - f(x)\| \le n\delta$.
3. F is smooth.

Let $\Omega = \frac{1}{2}B_X$, with μ the normalized Lebesgue measure on Ω:

$$\mu(A) = \frac{\operatorname{vol}(A \cap \left(\frac{1}{2}B_X\right))}{\operatorname{vol}(\frac{1}{2}B_X)}.$$

Define $T : X \to L_\infty(\mu, Z)$ as follows. For all $y \in X$,

$$(Ty)(x) = F'(x)(y) = \partial_y F(x).$$

This function is in L_∞ because

$$\|(Ty)(x)\| = \|F'(x)(y)\| \le \|F'(x)\|\,\|y\| \lesssim \|y\|.$$

This proves the upper bound.

We need

$$\frac{1}{D}\|y\| \lesssim \|Ty\|_{L_1(\mu,Z)} = \int_{\frac{1}{2}B_X} \|F'(x)y\|\,d\mu(x)$$
$$= \frac{1}{\mathrm{vol}(\frac{1}{2}B_X)}\int_{\frac{1}{2}B_X} \|F'(x)y\|\,dx.$$

We need two simple and general lemmas.

Lemma 3.5.5. *There exists an isometric embedding $J : X \to \ell_\infty$.*

This is a restatement of the Hahn-Banach theorem.

Proof. X^* is separable, so let $\{x_k^*\}_{k=1}^\infty$ be a set of linear functionals that is dense in $S_{X^*} = \partial B_{X^*}$. Let

$$J_X = (x_k^*(x))_{k=1}^\infty \in \ell_\infty.$$

For all x, $|x_k^*(x)| \le \|x_k^*\| \le \|x_k^*\|\,\|x\| = \|x\|$. By Hahn Banach, since every x admits a normalizing functional,

$$\|Jx\|_{\ell_\infty} = \sup_k |x_k^*(x)| = \sup_{x^*\in S_X} |x^*(x)| = \|x\|.$$

□

This may see a triviality, but there are 4 situations where this theorem will make its appearance.

Remark 3.5.6: Every separable metric space (X,d) admits an isometric embedding into ℓ_∞, given as follows. Take $\{x_k\}_{k=1}^\infty$ dense in X, and let

$$f(x) = (d(x,x_k) - d(x,x_0))_{k=1}^\infty.$$

Proof is left as an exercise.

Lemma 3.5.7 (Nonlinear Hahn-Banach Theorem)**.** *Let (X,d) be any metric space and $A \subseteq X$ any nonempty subset. If $f : A \to \mathbb{R}$ is L-Lipschitz then there exists $F : X \to \mathbb{R}$ that extends f and*

$$\|F\|_{Lip} = L = \|f\|_{Lip}.$$

The lemma says we can always extend the function and not lose anything in the Lipschitz constant. Recall the Hahn-Banach Theorem says that we can extend functionals from subspace and preserve the norm. Extension in vector spaces and extension in $\mathbb{R}$ are different.

This lemma seems general, but it is very useful.

We mimic the proof of the Hahn-Banach Theorem.

Proof. We will prove that one can extend f to one additional point while preserving the Lipschitz constant.

Then use Zorn's Lemma on the poset of all extensions to supersets ordered by inclusion (with consistency) to finish.

Let $A \subseteq X$, $x \in X \backslash A$. We define $t = F(x) \in \mathbb{R}$. To make F is Lipschitz, for all $a \in A$, we need

$$|t - f(a)| \le d_X(x, a).$$

Then we need

$$t \in \bigcap_{a \in A} [f(a) - d_X(x, a), f(a) + d_X(x, a)].$$

The existence of t is equivalent to

$$\bigcap_{a \in A} [f(a) - d_X(x, a), f(a) + d_X(x, a)].$$

By compactness it is enough to check this is true for finite intersections. Because we are on the real line (which is a total order), it is in fact enough to check this is true for pairwise intersections.

$$[f(a) - d_X(x, a), f(a) + d_X(x, a)] \cap [f(b) - d_X(x, b), f(b) + d_X(x, b)].$$

We need to check

$$\begin{aligned} & f(a) + d(x, a) \ge f(b) - d(x, b) \\ \iff & |f(b) - f(a)| \le d(x, a) + d(x, b). \end{aligned}$$

This is true by the triangle inequality:

$$|f(b) - f(a)| \le d(a, b) \le d(x, a) + d(x, b).$$

□

This theorem is used all over the place. Most of the time people just give a closed formula: define F on all the points at once by

$$F(x) = \inf \{f(a) + d(x, a) : a \in A\}.$$

Now just check that this satisfies Lipschitz. From our proof you ses exactly why they define F this way: it comes from taking the inf of the upper intervals (we can also take the sup of the lower intervals). The extension is not unique; there are many formulas for the extension.

There is a whole isometric theory about exact extensions: look at all the possible extensions, what is the best one? F is the pointwise smallest extension; the other one is the pointwise largest. The theory of absolutely minimizing extensions is related to an interesting nonlinear PDE called the infinite Laplacian. This is different from what we're doing because we lose constants in many places.

Corollary 3.5.8. *Let (X, d) be any metric space. Let $A \subseteq X$ be a nonempty subset $f : A \to \ell_\infty$ Lipschitz. Then there exists $F : X \to \ell_\infty$ that extends f and $\|F\|_{Lip} = \|f\|_{Lip}$.*

Proof. Note that saying $f : A \to \ell_\infty$, $f(a) = (f_1(a), f_2(a), \dots)$ has $\|f\|_{\text{Lip}} = 1$ means that f_i is 1-Lipschitz for every i. $\square$

f takes $\mathcal{N}_\delta \subseteq B_X$ to $f(\mathcal{N}_\delta) \subseteq Z$. f^{-1} takes it back to X; J is the embedding to ℓ_∞.

$$\begin{tikzcd}
B_X & Z \arrow[rrd, bend left, "G"] & & \\
\mathcal{N}_\delta \arrow[u, hook] \arrow[r, "f"] & f(\mathcal{N}_\delta) \arrow[u, hook] \arrow[r, "f^{-1}|_{f(\mathcal{N}_\delta)}"] \arrow[rr, bend right, "J\circ f^{-1}|_{f(\mathcal{N}_\delta)}"'] & X \arrow[r, "J"] & \ell_\infty
\end{tikzcd}$$

Note $J \circ f^{-1}|_{f(\mathcal{N}_\delta)}$ is D-Lipschitz. By nonlinear Hahn-Banach, there exists $G : Z \to \ell_\infty$ with $\|G\|_{\text{Lip}} \le D$ such that

$$G(f(x)) = J(x)$$

for all $x \in \mathcal{N}_\delta$.

We want G to be differentiable. We do this by convolving with a smooth bump function with small support to get $H : Z \to \ell_\infty$ such that

1. H is smooth
2. $\|H\|_{\text{Lip}} \le D$,
3. for all $x \in F(B_X)$, $\|G(x) - H(x)\| \le nD\delta$.

Define a linear operator

$$S : L_1(\mu, Z) \to \ell_\infty$$

by for all $h \in L_1(\mu, Z)$, $h : \frac{1}{2}B_X \to Z$,

$$Sh = \int_{\frac{1}{2}B_X} H'(F(x))(h(x))\, d\mu(x).$$

(Type checking: F is a point in X, we can differentiate at this point and get linear operator $Z \to \ell_\infty$. $h(x)$ is a point in Z, so we can plug into linear operator and get a point in ℓ_∞. This is a vector-valued integration; do the integration coordinatewise. So Sh is in ℓ_∞.)

Now

$$
\begin{aligned}
\|Sh\|_{\ell_\infty} &\le \int_{\frac{1}{2}B_X} \|H'(F(x))(h(x))\|_{\ell_\infty}\, d\mu(x)\\
&\le \int_{\frac{1}{2}B_X} \underbrace{\|H'(F(x))\|_{Z\to\ell_\infty}}_{=:D} \|h(x)\|\, d\mu(x)\\
&\le D\int_{\frac{1}{2}B_X} \|h(x)\|\\
&\le D\,\|h\|_{L_1(\mu,Z)}\\
\implies \|S\|_{L_1(\mu,Z)\to\ell_\infty} &\le D\\
S\underbrace{Ty}_{h} &= \int_{\frac{1}{2}B_X} H'(F(x))((Ty)(x))\,d\mu(x)\\
&= \int_{\frac{1}{2}B_X} H'(F(x))(F'(x)(y))\,d\mu(x)\\
&= \int_{\frac{1}{2}B_X} (H\circ F)'(x)(y)\,d\mu(x) && \text{chain rule}
\end{aligned}
$$

Now we show $H\circ F$ is very close to J; it's close to being invertible. (Recall $G(f(x)) = Jx$.)

This does not a priori mean the derivative is close. We use a geometric argument to say that if a function is sufficiently close to being an isometry, then the integral of its derivative on a finite dimensional space is close to being invertible. $H\circ F$ was a proxy to J.

Check that $H\circ F$ is close to J. For $y\in\mathcal{N}_\delta$, by choice of H,

$$
\begin{aligned}
\|H(F(y)) - Jy\|_{\ell_\infty} &\le \|H(F(y)) - G(F(y))\|_{\ell_\infty} + \|G(F(y)) - G(f(y))\|_{\ell_\infty}\\
&\le nD\delta + D\,\|F(y) - f(y)\|_Z\\
&\lesssim nD\delta.
\end{aligned}
$$

For general $x\in\frac{1}{2}B_X$, there exists $y\in\mathcal{N}_\delta$ such that $\|x-y\|_X \le 2\delta$. Then

$$
\begin{aligned}
\|H(F(x)) - Jx\|_{\ell_\infty} &\le nD\delta + D\delta + \delta\\
&\lesssim nD\delta.
\end{aligned}
$$

For all $x\in\frac{1}{2}B_X$, $\|H\circ F(x) - Jx\| \le CnD\delta$. Define $g(x) = H\circ F(x) - Jx$. Then $\|g\|_{L^\infty(\frac{1}{2}B_X)} \le CnD\delta$.

We need a geometric lemma.

Lemma 3.5.9. *Suppose $(V, \|\cdot\|_V)$ is any Banach space, $U = (\mathbb{R}^n, \|\cdot\|_U)$ is a n-dimensional Banach space. Let $g : B_U \to V$ be continuous on B_U and differentiable on int(B_U). Then*

$$
\left\|\frac{1}{vol(B_U)}\int_{B_U} g'(u)\,du\right\|_{U\to V} \le n\,\|g\|_{L_\infty(B_U)}.
$$

□

3/28: Finishing up Bourgain's Theorem

Previously we had X, Y Banach spaces with $\dim(X) = n$. Let $N_\delta \subseteq B_X$ be a δ-net, $\delta \le c/n^2 D$, $f : N_\delta \to Y$. We have

$$\frac{1}{D}\|x-y\| \le \|f(x) - f(y)\| \le \|x-y\|$$

WLOG $Y = \text{span}(f(N_\delta))$. The almost extension theorem gave us a smooth map from $F : X \to Y$, $\|F\|_{Lip} \lesssim 1$ and for all $x \in N_\delta$, $\|F(x) - f(x)\| \lesssim n\delta$. Letting μ be the normalized Lebesgue measure on $1/2B_X$, define $T : X \to L_\infty(\mu, Y)$ by

$$(Ty)(x) = F'(x)y$$

for all $y \in X$.

The goal was to prove

$$\|Ty\|_{L_1(\mu,Y)} \gtrsim \frac{1}{D}\|y\|$$

and the approach was to show for $y \in X$ the average $\frac{1}{\text{vol}(1/2B_X)}\int_{1/2B_X} \|F'(x)y\|dy$ is big.

Fix a linear isometric embedding $J : X \to l_\infty$., we proved that there exists $G : Y \to l_\infty$ such that $\forall x \in N_\delta, G(f(x)) = J(x)$. We have $\|G\|_{Lip} \le D$. Fix $H : Y \to l_\infty$ where H is smooth, $\|H\|_{Lip} \le D$, and $\forall x \in F(B_X)$, $\|H(x) - G(x)\| \le nD\delta$.

Define $S : L_1(\mu, Y) \to l_\infty$ by

$$Sh = \int_{1/2B_X} H'(F(x))(h(x))d\mu(x)$$

for all $h \in L_1(\mu, Y)$. Note that $\|S\|_{L_1(\mu,Y)\to l_\infty} \le D$. By the chain rule, $STy = \int_{1/2B_X}(H \circ F)'(x)yd\mu(x)$. We checked for every $x \in B_X$, $\|H(F(x)) - Jx\|_{l_\infty]} \lesssim nD\delta$.

Suppose you are on a point on the net. Then you know H is very close to G. F was $n\delta$ close to f, and H is D-Lipschitz, which is how we get $n\delta D$ (take a net point, find the closest point on the net, and use the fact that the functions are D-Lipschitz).

This is where we stopped. Now how do we finish?

We need a small geometric lemma:

Lemma 3.5.10. *Geometric Lemma.*
U, V are Banach spaces, $\dim(U) = n$. Let $U = (\mathbb{R}^n, \|\cdot\|_U)$. Suppose that $g : B_U \to V$ is continuous on B_U and smooth on the interior. Then

$$\left\|\frac{1}{\text{vol}(B_U)}\int_{B_U} g'(u)du\right\|_{U\to V} \le n\|g\|_{L_\infty(S_X)}$$

where the inside of the LHS norm should be thought of as an operator. Let R be the normalized integral operator. Then

$$Rx = \frac{1}{\text{vol}(B_U)}\int_{B_U} g'(u)xdu$$

The way to think of using this is if you have an L_∞ bound, you get an L_1 bound. Assuming this lemma, let's apply it to $g = H \circ F - J$. We get

$$\left\| \int_{1/2B_X} ((H\circ F)' - J)\, d\mu(x) \right\|_{X\to l_\infty} \lesssim n^2 D\delta$$
$$\left\| \int_{1/2B_X} (H\circ F)'(x) d\mu(x) - J \right\|_{X\to l_\infty} = \|ST - J\|_{X\to l_\infty} \lesssim n^2 D\delta$$

Now we want to bound from below $\|Ty\|_{L_1(\mu,Y)}$. We have

$$\begin{aligned}\|Ty\|_{L_1(\mu,Y)} &\geq \frac{\|STy\|_{l_\infty}}{\|S\|_{L_1(\mu,Y)\to l_\infty}}\\ &\geq \frac{1}{D}\|STy\|_{l_\infty} \geq \frac{1}{D}\left(\|Jy\| - \|ST-J\|_{X\to l_\infty}\|y\|\right)\\ &= \frac{1}{D}\left(\|y\| - (n^2 D\delta\|y\|)\right)\end{aligned}$$

There is a bit of magic in the punchline. We want to bound the operator below, and we understand it as an average of derivatives. We succeeded to show the function itself is small, and there is our geometric lemma which gives a bound on the derivative if you know a bound on the function.

Now let's prove the lemma.

Proof of Lemma 3.5.10. Fix a direction $y \in \mathbb{R}^n$ and normalize so that $\|y\|_2 = 1$. For every $u \in \mathrm{Proj}_{y^\perp}(B_U)$, let $a_U \leq b_U \in \mathbb{R}$ be such that $u + \mathbb{R}y \cap B_U = u + [a_U, b_U]y$ (basically, this is the intersection of the projection line with the ball).

Using Fubini,

$$\begin{aligned}\left\|\frac{1}{\mathrm{vol}(B_U)}\int_{B_U} g'(u)du\right\|_V &= \left\|\frac{1}{\mathrm{vol}(B_U)}\int_{\mathrm{Proj}_{y^\perp}(B_U)}\int_{a_U}^{b_U}\frac{d}{ds}g(u+sy)dsdu\right\|_V\\ &= \left\|\frac{1}{\mathrm{vol}(B_U)}\int_{\mathrm{Proj}_{y^\perp}(B_U)}\left(g(u+b_Uy) - g(u+a_Uy)\right)du\right\|_V\\ &\leq \frac{1}{\mathrm{vol}_n(B_U)}\cdot 2\mathrm{vol}_{n-1}(\mathrm{Proj}_{y^\perp}(B_U))\|g\|_{L_\infty(S_X,V)}\end{aligned}$$

We need to show that

$$\frac{2\mathrm{vol}_{n-1}\left(\mathrm{Proj}_{y^\perp}(B_U)\right)}{\mathrm{vol}_n(B_U)} \leq n\|y\|_U$$

The convex hull of y over the projection is the cone over the projection.

$$\mathrm{vol}_n(\mathrm{conv}\left(\frac{y}{\|y\|_U} \cup \mathrm{Proj}_{y^\perp}(B_U)\right)) = \frac{1}{n\|y\|_U}\cdot \mathrm{vol}_{n-1}\left(\mathrm{Proj}_{y^\perp}(B_U)\right)$$

Letting $K = \mathrm{conv}\left(\left\{\pm\frac{y}{\|y\|_U}\right\} \cap \mathrm{Proj}_{y^\perp}(B_U)\right)$, this is the same as saying that $\mathrm{vol}_n(K) \leq \mathrm{vol}_n(B_U)$. Note that K is the double cone, that is where the factor of 2 got absorbed. This is an application of Fubini's theorem.

Geometrically, the idea is that when we look on the projective lines for whatever part of the cone is not in our ball set, we will be able to fit the outside inside the set. In formulas, c_U is the largest multiple of u which is inside the boundary of the cone. $c_U \geq 1$.

$$K = \bigcup_{u \in \text{Proj}_{y^\perp}(B_U)} \left(u + \left[-\frac{c_U - 1}{c_U \|y\|_U}, +\frac{c_U - 1}{c_U \|y\|_U} \right] \right)$$

We also have

$$\frac{1}{c_U}\left(c_U u + a_{c_U u} y\right) \pm \left(1 - \frac{1}{c_U}\right) \frac{y}{\|y\|_U} \in B_U$$

and we get that $K \subset B_U$ by Fubini. □

Thus, we've completed the proof of Bourgain's theorem for the semester. The big remaining question is can we do something like this in the general case?

6 Kirszbraun's Extension Theorem

Last time we proved nonlinear Hahn-Banach theorem. I want to prove one more Lipschitz Extension theorem, which I can do in ten minutes which we will need later in the semester.

Theorem 3.6.1 (Kirszbraun's extension theorem (1934))**.** *Let H_1, H_2 be Hilbert spaces and $A \subseteq H_1$. Let $f : A \to H_2$ Lipschitz. Then, there exists a function $F : H_1 \to H_2$ that extends f and has the same Lipschitz constant ($\|F\|_{Lip} = \|f\|_{Lip}$).*

We did this for real valued functions and l_∞ functions. This version is non-trivial, and relates to many open problems.

Proof. There is an equivalent geometric formulation. Let H_1, H_2 be Hilbert spaces $\{x_i\}_{i \in I} \subseteq H_1$ and $\{y_i\}_{i \in I} \subseteq H_2$, $\{r_i\}_{i \in I} \subseteq (0, \infty)$. Suppose that $\forall i, j \in I$, $\|y_i - y_j\|_{H_2} \leq \|x_i - y_i\|_{H_1}$. If

$$\bigcap_{i \in I} B_{H_1}(x_i, r_i) \neq \emptyset$$

then

$$\bigcap_{i \in I} B_{H_1}(y_i, r_i) \neq \emptyset$$

as well.

Intuitively, this says the configuration of points in H_2 are a squeezed version of the H_1 points. Then, we're just saying something obvious. If there's some point that intersects all balls in H_1, then using the same radii in the squeezed version will also be nonempty. Our geometric formulation will imply extension. We have $f : A \to H_2$, WLOG $\|f\|_{Lip} = 1$. For all $a, b \in A$, $\|f(a) - f(b)\|_{H_2} \leq \|a - b\|_{H_1}$. Fix any $x \in H_1 \setminus A$. What can we say about the intersection of the following balls?: $\bigcap_{a \in A} B_{H_1}(a, \|a - x\|_{H_1})$. Well by design it is not empty since x is in this set. So the conclusion from the geometric formulation says

$$\bigcap_{a \in A} B_{H_2}\left(f(a), \|a - x\|_{H_1}\right) \neq \emptyset$$

So take some y in this set. Then $\|y - f(a)\|_{H_2} \le \|x - a\|_{H_2}$ $\forall a \in A$. Define $F(x) = y$. Then we can just do the one-more-point argument with Zorn's lemma to finish.

Let us now prove the geometric formulation. It is enough to prove the geometric formulation when $|I| < \infty$ and H_1, H_2 are finite dimensional. To show all the balls intersect, it is enough to show that finitely many of them intersect (this is the finite intersection property: balls are weakly compact). Now the minute I is finite, we can write $I = \{1, \cdots, n\}$, $H_1' = \text{span}\{x_1, \cdots, x_n\}, H_2' = \text{span}\{y_1, \cdots, y_n\}$. This reduces everything to finite dimensions. We have a nice argument using Hahn-Banach. FINISH THIS □

Remark 3.6.2: In other norms, the geometric formulation in the previous proof is just not true. This effectively characterizes Hilbert spaces. Related is the Kneser-Poulsen conjecture, which effectively says the same thing in terms of volumes:

Conjecture 3.6.3 (Kneser-Poulsen)**.** *Let $x_1, \ldots, x_k; y_1, \ldots, y_k \in \mathbb{R}^n$ and $\|y_i - y_j\| \le \|x_i - x_j\|$ for all i, j. Then $\forall r_i \ge 0$*

$$\text{vol}\left(\bigcap_{i=1}^k B(y_i, r_i)\right) \ge \text{vol}\left(\bigcap_{i=1}^k B(x_i, r_i)\right)$$

This is known for $n = 2$, where volume is area using trigonometry and all kinds of ad-hoc arguments. It's also known for $k = n + 2$.

3/30/16

Here is an equivalent formulation of the Kirszbraun extension theorem.

Theorem 3.6.4. *Let H_1, H_2 be Hilbert spaces, $\{x_i\}_{i \in I} \subseteq H_1$, $\{y_i\}_{i\in I} \subseteq H_2$, $\{r_i\}_{i=1}^\infty \subseteq (0, \infty)$. Suppose that $\|y_i - y_j\|_{H_2} \le \|x_i - x_j\|_{H_2}$ for all $i, j \in I$. Then $\bigcap_{i\in I} B_{H_1}(x_i, r_i) \ne \phi$ implies $\bigcap_{i\in I} B_{H_2}(y_i, r_i) \ne \phi$.*

Proof. By weak compactness and orthogonal projection, it is enough to prove this when $I = \{1, \ldots, n\}$ and H_1, H_2 are both finite dimensional. Fix any $x \in \bigcap_{i=1}^n B_{H_1}(x_i, r_i)$. If $x = x_{i_0}$ for some $i_0 \in \{1, \ldots, n\}$, $\|y_{i_0} - y_i\|_{H_2} \le \|x_{i_0} - x_i\|_{H_1} \le r_i$, and $y_{i_0} \in \bigcap_{i=1}^n B_{H_2}(y_i, r_i)$. Assume $x \notin \{x_1, \ldots, x_n\}$.

Define $f : H \to \mathbb{R}$ by

$$f(y) = \max\left\{\frac{\|y - y_1\|_{H_2}}{\|x - x_1\|_{H_1}}, \ldots, \frac{\|y - y_n\|_{H_2}}{\|x - x_n\|_{H_2}}\right\}.$$

Note that f is continuous and $\lim_{\|y\|_{H_2} \to \infty} f(y) = \infty$.

So, the minimum of $f(y)$ over $\mathbb{R}^n$ is attained. Let

$$m = \min_{y \in \mathbb{R}^n} f(y).$$

Fix $y \in \mathbb{R}^n$, $f(y) = m$.

Observe that we are done if $m \le 1$. Suppose by way of contradiction $m > 1$. Let J be the indices where the ratio is the minimum.

$$J = \left\{ i \in \{1, \dots, n\} : \frac{\|y - y_i\|_{H_2}}{\|x - x_i\|_{H_1}} = m \right\}.$$

By definition, if $j \in J$, then $\|y - y_j\|_{H_2} = m \|x - x_j\|_{H_1}$, and for $j \notin J$, $\|y - y_j\|_{H_2} < m \|x - x_j\|_{H_1}$.

Claim 3.6.5. $y \in \operatorname{conv}(\{y_j\}_{j \in J})$.

Proof. If not, find a separating hyperplane. If we move y a small enough distance towards $\operatorname{conv}(\{y_j\}_{j\in J})$ perpendicular to the separating hyperplane to y', then for $j \in J$,

$$\|y' - y_j\|_{H_2} < \|y - y_j\|_{H_2} = m \|x - x_j\|_{H_1}.$$

For $j \in J$, it is still true that $\|y' - y_j\|_{H_2} < m \|x - x_j\|_{H_1}$. Then $f(y') < m$, contradicting that y is a minimizer. □

By the claim, there exists $\{\lambda_j\}_{j \in J}$, $\lambda_j \ge 0$, $\sum_{j \in J} \lambda_j = 1$, $y = \sum_{j\in J} \lambda_j y_j$.

We use this the coefficients of this convex combination to define a probability distribution. Define a random vector in H_1 by X, with

$$\mathbb{P}(X = x_j) = \lambda_j, \quad j \in J.$$

For all $j \in \{1, \dots, n\}$, let $y_j = h(x_j)$. Let $\mathbb{E}[h(X)] = \sum_{j \in J} \lambda_j y_j = y$. Let X' be an independent copy of X.

Using $\|h(X) - y\|_{H_2} = m \|X - x\|_{H_1}$, we have

$$\begin{aligned}
\mathbb{E}\|h(X) - \mathbb{E}h(X)\|_{H_2}^2 &= \mathbb{E}\|h(X) - y\|_{H_2}^2 && (3.25)\\
&= m^2 \mathbb{E}\|X - x\|_{H_1}^2 && (3.26)\\
&> \mathbb{E}\|X - x\|_{H_1}^2 && (3.27)\\
&\ge \mathbb{E}\|X - EX\|_{H_1}^2 && (3.28)
\end{aligned}$$

In Hilbert space, it is always true that

$$\mathbb{E}\|X - EX\|_{H_1}^2 = \frac{1}{2}\mathbb{E}\|X - X'\|_{H_2}^2.$$

Using this on both sides of the above,

$$\frac{1}{2}\mathbb{E}\|h(X) - h(X')\|_{H_2}^2 > \frac{1}{2}\|X - X'\|_{H_1}^2.$$

So far we haven't used the only assumption on the points. By the assumption $\|X - X'\|_{H_1} \ge \|h(X) - h(X')\|_{H_2}$. This is a contradiction. □

This is not true in L^p spaces when $p \neq 2$, but there are other theorems one can formulate (ex. with different radii). We have to ask what happens to each of the inequalities in the proof. As stated the theorem is a very "Hilbertian" phenomenon.

Definition 3.6.6: Let $(X, \|\cdot\|_X)$ be a Banach space and $p \geq 1$. We say that X has **Rademacher type** p if for every n, for every $x_1, \ldots, x_n \in X$,

$$\left(\mathbb{E}_{\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)\in\{\pm1\}^n} \left[\left\| \sum_{i=1}^n \varepsilon_i X_i \right\|_X^p \right] \right)^{\frac{1}{p}} \leq T \left(\sum_{i=1}^n \|X_i\|_X^p \right)^{\frac{1}{p}}.$$

The smallest T is denoted $T_p(X)$.

Definition (Definition 1.1.13)**:** A metric space (X, d_X) is said to have **Enflo type** p if for all $f : \{\pm1\}^n \to X$,

$$\left(\mathbb{E}_{\varepsilon}[d_X(f(\varepsilon), f(-\varepsilon))^p] \right)^{\frac{1}{p}} \lesssim_X \left(\sum_{j=1}^n \mathbb{E}_{\varepsilon}[d(f(\varepsilon), f(\varepsilon_1, \ldots, \varepsilon_{j-1}, -\varepsilon_j, \varepsilon_{j+1}, \cdots))^p] \right)^{\frac{1}{p}}.$$

If X is also a Banach space of Enflo type p, then it is also of Rademacher type p.

Question 3.6.7: Let X be a Banach space of Rademacher type p. Does X also have Enflo type p?

We know special Banach spaces for which this is true, like L^p spaces, but we don't know the anser in full generality.

Theorem 3.6.8 (Pisier)**.** thm:pisier *Let X be a Banach space of Rademacher type p. Then X has Enflo type q for every $1 < q < p$.*

We first need the following.

Theorem 3.6.9 (Kahane's Inequality)**.** *For every $\infty > p, q \geq 1$, there exists $K_{p,q}$ such that for every Banach space, $(X, \|\cdot\|_X)$ for every $x_1, \ldots, x_n \in X$,*

$$\left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_X^p \right)^{\frac{1}{p}} \leq K_{p,q} \left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_X^q \right)^{\frac{1}{q}}.$$

In the real case this is Khintchine's inequality.

By Kahane's inequality, X has type p iff

$$\left(\mathbb{E} \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|_X^q \right)^{\frac{1}{q}} \lesssim_{X,p,q} \left(\sum_{i=1}^n \|x_i\|_X^p \right)^{\frac{1}{p}}.$$

Define $T : \ell_p^n(X) \to L_p(\{\pm1\}^n, X)$ by $T(x_1, \ldots, x_n)(\varepsilon) = \sum_{i=1}^n \varepsilon_i X_i$.

(Here $\|f\|_{L_p(\{\pm 1\}^n,X)} = \left(\frac{1}{2^n}\sum_{\varepsilon\in\{\pm 1\}^n}\|f(\varepsilon)\|_X^p\right)^{\frac{1}{p}}$.)

Rademacher type p mens

$$\|T\|_{\ell_p^n(X)\to L_q(\{\pm 1\}^n,X)} \lesssim_{X,p,q} 1.$$

For an operator $T : Y \to Z$, the adjoint $T^* : Z^* \to Y^*$ satisfies

$$\|T\|_{Y\to Z} = \|T^*\|_{Z^*\to Y^*}.$$

Let $p^* = \frac{p}{p-1}$ be the dual of p ($\frac{1}{p}+\frac{1}{p^*}=1$). Now

$$L_q(\{\pm 1\}^n, X)^* = L_{q^*}(\{\pm 1\}^n, X^*).$$

Here, $g^* : \{\pm 1\}^n \to X^*$, $f : \{\pm 1\}^n \to X$, $g^*(f) = \mathbb{E}g^*(\varepsilon)(f(\varepsilon))$, $\ell_p^n(X) = \ell_{p^*}^n(X^*)$.

Note

$$\|T^*\|_{L_{q^*}(\{\pm 1\}^n,X^*)\to \ell_{p^*}^n(X^*)} \lesssim 1,$$

For $T : Y \to Z$, $T^* : Z^* \to Y^*$ is defined by $T^*(z^*)(y) = z^*(Ty)$.

For $g^* : \{\pm 1\}^n \to X^*$, $g^* \sum_{A\subseteq\{1,\ldots,n\}} \widehat{g}^*(A)W_A$. We claim

$$T^*g^* = (\widehat{g}^*(\{1\}),\ldots,\widehat{g}(\{n\})).$$

We check

$$\begin{aligned}
T^*g^*(x_1,\ldots,x_n) &= g^*(T(x_1,\ldots,x_n)) \\
&\le \mathbb{E}g^*(\varepsilon)\left(\sum_{i=1}^n \varepsilon_i X_i\right) \\
&= \sum_{i=1}^n (\mathbb{E}g^*(\varepsilon)\varepsilon_i)(X_i) \\
&= \sum_{i=1}^n \left(\sum_{A\subseteq\{1,\ldots,n\}} \widehat{g}^*(A)(\mathbb{E}W_A(\varepsilon)\varepsilon_i)\right)(x_i) \\
&= \sum_{i=1}^n \widehat{g}^*(\{i\})(x) \\
&= (\widehat{g}^*(\{1\}),\ldots,\widehat{g}^*(\{n\}))(x_1,\ldots,x_n).
\end{aligned}$$

Rademacher type p means for all $g_i^* : \{\pm 1\}^n \to X^*$, for all $q \in [1,\infty)$,

$$\left(\sum_{i=1}^n \|\widehat{g}^*(\{i\})\|_X^{p^*}\right) \lesssim \|g^*\|_{L_q(\{\pm 1\}^n,X^*)}$$

Theorem 3.6.10 (Pisier)**.** *If X has Rademacher type p and $1 \le a < p$, then for every $b > 1$, for all $f : \{\pm 1\}^n \to X$ with $\mathbb{E}f = 0$,*

$$\|f\|_b \precsim \left\|\left(\sum_{j=1}^n \|\partial_j f\|_X^a\right)^{\frac{1}{a}}\right\|_b.$$

Here, for $f : \{\pm 1\}^n \to X$, $\partial_j f : \{\pm 1\}^n \to X$ is defined by

$$\partial_j f(\varepsilon) = \frac{f(\varepsilon) - f(\varepsilon_1, \dots, \varepsilon_{j-1}, -\varepsilon_j, \varepsilon_{j+1}, \dots \varepsilon_n)}{2} = \sum_{\substack{A \subseteq \{1,\dots,n\} \\ j \in A}} \widehat{f}(A) W_A.$$

Note $\partial_j^2 = \partial_j$.

The Laplacian is $\Delta = \sum_{j=1}^n \partial_j = \sum_{j=1}^n \partial_j^2$. We've proved that

$$\Delta f = \sum_{A \subseteq \{1,\dots,n\}} |A| \widehat{f}(A) W_A.$$

The heat semigroup on the cube[5] is for $t > 0$,

$$e^{-t\Delta} f = \sum_{A \subseteq \{1,\dots,n\}} e^{-t|A|} \widehat{f}(A) W_A.$$

This is a function on the cube

$$\varepsilon \mapsto \left(\sum_{j=1}^n \|\partial_j f(\varepsilon)\|_X^a \right)^{\frac{1}{a}} \in [0, \infty).$$

For $b = a$,

$$\|f\|_a \lesssim \left(\sum_{j=1}^n \|\partial_j f\|_{L_a(\{\pm 1\}^n, X)}^q \right)^{\frac{1}{a}}.$$

Let $f : \{\pm 1\}^n \to X$ be

$$\begin{aligned} \|f(\varepsilon) - f(-\varepsilon)\|_a &\le \|f(\varepsilon) - \mathbb{E} f\|_a + \mathbb{E} \|\mathbb{E} f - \mathbb{E} f(-\varepsilon)\|_a \\ &\lesssim \left(\sum_{j=1}^n \|f(\varepsilon) - f(\varepsilon_1, \dots, -\varepsilon_j, \dots, \varepsilon_n)\|_X^a \right)^{\frac{1}{a}}. \end{aligned}$$

We need some facts about the heat semigroup.

Fact 3.6.11 (Fact 1)**:** Heat flow is a contraction:

$$\left\| e^{-t\Delta} f \right\|_{L_q(\{\pm 1\}^n, X)} \le \|f\|_{L_q(\{\pm 1\}^n, X)}.$$

Proof. We have

$$e^{-t\Delta} f = \sum_{A \subseteq \{1,\dots,n\}} e^{-t|A|} \widehat{f}(A) W_A = \sum R_t * f$$

[5] also known as the noise operator

Here the convolution is defined as $\frac{1}{2^n}\sum_{\delta\in\{\pm1\}^n} R_t(\varepsilon\delta)f(\delta)$. This is a vector valued function. In the real case it's Parseval, multiplying the Fourier coefficients. It's enough to prove for the real line. Identities on the level of vectors. Here

$$R_t(\varepsilon) = \sum_{A\subseteq\{1,\ldots,n\}} e^{-t|A|}W_A(\varepsilon);$$

this is the Riesz product.

The function is

$$= \prod_{i=1}^n (1+e^{-t}\varepsilon_i) = 0, t>0.$$

$\mathbb{E}R_t = 1$ so R_t is a probability measure and heat flow is an averaging operator. $\square$

4/11: Continue Theorem 3.6.8. The dual to having Rademacher type p: for all $g^*: \{\pm1\}^n \to X^*$, $\left(\sum_{i=1}^n \|\widehat{g}(\{i\})\|_{X^*}^{p^*}\right)^{\frac{1}{p^*}} \lesssim \|g^*\|_{L^{r^*}(\{\pm1\}^n,X^*)}$ for every $1<r<\infty$.

Note that for the purpose of proving Theorem 3.6.8 we only need $r=p$ in Kahane's inequality. However, I recommend that you read up on the full inequality.

For $f:\{\pm1\}^n\to X$, define $\partial_j f:\{\pm1\}^n \to X$ by $\partial_j f = \frac{f(\varepsilon)-f(\varepsilon_1,\ldots,-\varepsilon_j,\ldots,\varepsilon_n)}{2}$. Then $\partial_j^2 = \partial_j = \partial_j^*$ and $\partial_j W_A = \begin{cases} W_A, & j\in A\\ 0, & j\notin A.\end{cases}$

The Laplacian is $\Delta = \sum_{j=1}^n \partial_j = \sum_{j=1}^n \partial_j^2$,

$$\Delta f = \sum_{A\subseteq[n]} |A|\widehat{f}(A)W_A.$$

If $t>0$ then

$$e^{-t\Delta} = \sum_{A\subseteq[n]} e^{-t|A|}W_A$$

is the heat semigroup. It is a contraction. $e^{-t\Delta} = \left(\prod_{i=1}^n(1+e^{-t}\varepsilon_i)\right) * f$.

We have

$$\left\|e^{-t\Delta}f\right\|_{L_p(\{\pm1\}^n,X)} \ge e^{-nt}\|f\|_{L^p(\{\pm1\}^n,X)} \tag{3.29}$$

$$e^{-t\Delta}(V_{[n]}e^{-t\Delta}f) = e^{-tn}W_{[n]}f \tag{3.30}$$

where $W_{[n]}(\varepsilon) = \prod_{i=1}^n \varepsilon_i$. For $f = \sum_{A\subseteq[n]} \widehat{f}(A)W_A$, we have

$$\begin{aligned}
W_{[n]}e^{-t\Delta}f &= \sum_{A\subseteq[n]} e^{-t|A|}\widehat{f}(A)W_{[n]\backslash A} && (3.31)\\
e^{-t\Delta}(W_{[n]}e^{-t\Delta}f) &= \sum_{A\subseteq[n]} e^{-t|A|}\widehat{f}(A)e^{-t(n-|A|)} && (3.32)\\
&= e^{-tn}\sum_{A\subseteq[n]} \widehat{f}(A)W_{[n]\backslash A} && (3.33)\\
&= e^{-tn}W_{[n]}f && (3.34)\\
e^{-tn}\left\|W_{[n]}f\right\|_{L_p(\{\pm1\}^n,X)} &= \left\|e^{-t\Delta}(W_{[n]}(e^{-t\Delta}f))\right\|_p && (3.35)\\
&\le \left\|\cancel{W_{[n]}}e^{-t\Delta}f\right\|_p. && (3.36)
\end{aligned}$$

The key claim is the following.

Claim 3.6.12. clm:pis1 *Suppose X has Rademacher type p. For all $1<r<\infty$, for all $t>0$, for all $g^*:\{\pm1\}^n\to X^*$,*

$$\left\|\left(\left\|e^{-t\Delta}\partial_j g^*(\varepsilon)\right\|_{X^*}^{q^*}\right)^{\frac{1}{q^*}}\right\|_{L_{r^*}(\{\pm1\}^n)} \lesssim_{X,r,p} \frac{\|g^*\|_{L_{r^*}(\{\pm1\}^n,X^*)}}{(e^t-1)^{\frac{p^*}{q^*}}}.$$

Assume the claim. We prove the following.

Claim 3.6.13. clm:pis2 *If X has Rademacher type p, $1<q<p$, and $1<r<\infty$, then for every function $f:\{\pm1\}^n\to X$ with $\mathbb{E}f = 0 = \widehat{f}(\phi)$ we have*

$$\|f\|_{L_r(\{\pm1\}^n,X)} \lesssim \frac{1}{p-q}\left\|\left(\sum_{j=1}^n \|\partial_j f(\varepsilon)\|_X^q\right)^{\frac1q}\right\|_{L_r(\{\pm1\}^n,X)}.$$

When $r=q$,

$$\begin{aligned}
\|f-\mathbb{E}f\|_q &\lesssim \left(\frac{1}{p-q}\right)^q \mathop{\mathbb{E}}_{\varepsilon}\sum_{j=1}^n \|\partial_j f(\varepsilon)\|_X^q && (3.37)\\
&= \left(\frac{1}{p-q}\right)^q \frac{1}{2^q}\sum_{j=1}^n \|f(\varepsilon)-f(\varepsilon_1,\ldots,-\varepsilon_j,\varepsilon_n)\|_X^q && (3.38)\\
\|f(\varepsilon)-f(-\varepsilon)\|_q &\le && (3.39)\\
&= 2\|f(\varepsilon)-\mathbb{E}f\|_q && (3.40)\\
&\lesssim \frac{1}{p-q}\left(\sum_{j=1}^n \mathbb{E}\|f(\varepsilon)-f(\varepsilon_1,\ldots,-\varepsilon_j,\varepsilon_n)\|^2\right) && (3.41)
\end{aligned}$$

by definition of Enflo type.

We show that the Key Claim 3.6.12 implies Claim 3.6.13.

Proof. Normalize so that

$$\left\| \left(\sum_{j=1}^n \|\partial_j f(\varepsilon)\|_X^q \right)^{\frac{1}{q}} \right\|_{L_r(\{\pm 1\}^n, X)} = 1.$$

For every $s > 0$, by Hahn-Banach, take $g_s^* : \{\pm 1\}^n \to X^*$ such that $\|g_s^*\|_{L_{r^*}(\{\pm 1\}^n, X^*)} = 1$ and

$$\mathbb{E}\left[g_s^*(\varepsilon)(\Delta e^{-s\Delta} f(\varepsilon)) \right] = \left\| \Delta e^{-s\Delta} f \right\|_{L_r(\{\pm 1\}^n, X)}.$$

Write this as $\left\langle g_s^*, \Delta e^{-s\Delta} f \right\rangle$.

Recall that $L_r(\{\pm 1\}^n, X)^* = L_{r^*}(\{\pm 1\}^n, X^*)$. Taking $h \in L_r(\{\pm 1\}^n, X)^*$ and $g^* \in L_{r^*}(\{\pm 1\}^n, X^*)$, we have

$$g^*(h) = \langle g^*, h \rangle = \mathbb{E}[g^*(\varepsilon) h(\varepsilon)] = \mathbb{E}[\langle g^*, h \rangle (\varepsilon)].$$

We have

$$\Delta e^{-s\Delta} f = \sum_{A \subseteq [n]} |A| e^{-s|A|} \widehat{f}(A) W_A \tag{3.42}$$

$$\left\| \Delta e^{-s\Delta} f \right\|_{L^r(X)} = \left\langle g_s^*, \sum \partial_j^2 e^{-s\Delta} f \right\rangle \tag{3.43}$$

$$= \sum_{j=1}^n \left\langle g_s^*, \partial_j e^{-s\Delta} \partial_j f \right\rangle \tag{3.44}$$

$$= \sum_{j=1}^n \left\langle e^{-s\Delta} \partial_j g_s^*, \partial_j f \right\rangle \tag{3.45}$$

$$= \underbrace{\left(e^{-s\Delta} \partial_j g_s^* \right)_{j=1}^n}_{\in L_{r^*}(\ell_{q^*}^n(X^*))} \underbrace{(\partial_j f)_{j=1}^n}_{\in L_r(\ell_q^n(X))}. \tag{3.46}$$

Note that $(L_r(\ell_q^n(X)))^* = L_{r^*}(\ell_{q^*}^n(X^*))$, so we have a pairing here.

$$\leq \left\| \left(\sum_{j=1}^n \left\| e^{-s\Delta} \partial_j g_s^*(\varepsilon) \right\|_{X^*}^{q^*} \right)^{\frac{1}{q}} \right\|_{L^{r^*}(X^n)} \left\| \left(\sum_{j=1}^n \|\partial_j f(\varepsilon)\|_X^q \right)^{\frac{1}{q}} \right\|_{L^r(X)} \tag{3.47}$$

$$\leq \frac{\|g_s^*\|_{L_{r^*}(X^*)}}{(e^t - 1)^{\frac{p^*}{q^*}}} \qquad \text{by Key Claim 3.6.12} \tag{3.48}$$

$$\lesssim \frac{1}{(e^t - 1)^{\frac{p^*}{q^*}}} \cdot \left\| \Delta e^{-s\Delta} f \right\|_{L_r(X)} \lesssim_{X,p,r} \frac{1}{(e^t - 1)^{\frac{p^*}{q^*}}}. \tag{3.49}$$

Integrating,

$$\int_0^\infty \Delta e^{-s\Delta} f = \sum_{A \subseteq [n]} \int_0^\infty |A| e^{-s|A|} \widehat{f}(A) W_A. \tag{3.50}$$

Then

$$\|f\|_{L_r(X)} = \left\|\int_0^\infty \Delta e^{-s\Delta} f\, ds\right\|_{L_r(X)} \tag{3.51}$$
$$\leq \int_0^\infty \left\|\Delta e^{-s\Delta} f\right\|_{L^r(X)} ds \tag{3.52}$$
$$\leq \int_0^\infty \frac{ds}{(e^t-1)^{\frac{p^*}{q^*}}} < \infty. \tag{3.53}$$

□

We now prove the key claim 3.6.12.

Proof of Key Claim 3.6.12. This is an interpolation argument. We first introduce a natural operation.

Given $g^* : \{\pm 1\}^n \to X^*$, for every $t > 0$ define a mapping $g_t^* : \{\pm 1\}^n \times \{\pm 1\}^n \to X^*$ as follows. For $\delta \in \{\pm 1\}^n$,

$$g^*(\varepsilon, \delta) = \sum_{A \subseteq [n]} \widehat{g}^*(A) \prod_{i \in A} (e^{-t}\varepsilon_i + (1-e^{-t})\delta_i) \tag{3.54}$$
$$= g^*(e^{-t}\varepsilon + (1-e^{-t})\delta) \tag{3.55}$$

Think of $g^*(\varepsilon) = \sum_{A \subseteq [n]} \widehat{g}^*(A) \prod_{i\in A} \varepsilon_i$ (extended outside $\{\pm 1\}^n$ by interpolation).

Observe

$$g^*(\varepsilon, \delta) = \sum_{B \subseteq [n]} e^{-t|B|}(1-e^{-t})^{n-|B|} g^*\left(\sum_{i\in B} \varepsilon_i e_i + \sum_{i \notin B} \delta_i e_i\right) \tag{3.56}$$

$$g^*\left(\sum_{i\in B} \varepsilon_i e_i + \sum_{i \notin B} \delta_i e_i\right) = \sum_{A \subseteq [n]} \widehat{g^*}(A) W(A \cap B)(\varepsilon) W_{A \setminus B}(\delta) \tag{3.57}$$

$$\sum_{B \subseteq [n]} e^{-t|B|}(1-e^{-t})^{n-|B|} g^*(\sum_{i\in B} \varepsilon_i e_i + \sum_{i \notin B} \delta_i e_i) \tag{3.58}$$
$$= \sum_{B \subseteq [n]} e^{-t|B|}(1-e^{-t})^{n-|B|} \sum_{A \subseteq [n]} \widehat{g^*}(A) W_{A\cap B}(\varepsilon) W_{A \setminus B}(\delta) \tag{3.59}$$
$$= \sum_{A \subseteq [n]} \widehat{g^*}(A) \sum_{B \subseteq [n]} e^{-t|B|}(1-e^{-t})^{n-|B|} W_{A\cap B}(\varepsilon) W_{A \setminus B}(\delta). \tag{3.60}$$

Let $\gamma = \begin{cases} \varepsilon_i, & \text{w.p. } e^{-t} \\ \delta_i, & \text{w.p. } 1-e^{-t}. \end{cases}$. Then

$$t\left(\prod_{i\in A} \gamma_i\right) = \prod_{i \in A} (e^{-t} + (1-e^{-t})\delta_i \tag{3.61}$$

$$\sum_A \widehat{g^*}(A) \mathbb{E} w_A(\gamma) = \mathbb{E}\left(\sum_A \widehat{g^*}(A) W_A(\gamma)\right) \tag{3.62}$$
$$= \sum_{B \subseteq [n]} e^{-t|B|}(1-e^{-t})^{n-|B|} g^*(\gamma) \tag{3.63}$$

Instead of a linear interpolation, we did a random interpolation.

What is $\|g_t^*\|_{L_{r^*}(\{\pm 1\}^n\times\{\pm 1\}^n, X^*)} \le \sum_{B\subseteq[n]} e^{-t|B|(1-e^{-t})^{n-|B|}}$? $\square$

4/13: Continue Theorem 3.6.8.

We're proving a famous and nontrivial theorem, so there are more computations (this is the only proof that's known).

We previously reduced everything to the following "Key Claim" 3.6.12.

Fix $t > 0$. Define $g_t^* : \{\pm 1\}^n \times \{\pm 1\}^n \to X^*$. Then

$$\begin{aligned} g_t^*(\varepsilon,\delta) &= g^*(e^{-t}\varepsilon + (1-e^{-t})\delta) \\ &= \sum_{A\subseteq\{1,\cdots,n\}} \hat{g}^*(A)\prod_{i\in A}\left(e^{-t}\varepsilon_i + (1-e^{-t})\delta_i\right) \end{aligned}$$

Then it's a fact that $\|g_t^*\|_{L_{r^*}(\{\pm 1\}^n\times\{\pm 1\}^n, X^*)\le\|g^*\|_{L_r(\{\pm 1\}^n, X^*)}}$ which is true for any Banach space from a Bernoulli interpretation.

Let us examine the Fourier expansion of the original definition of g^* in the variable δ_i. What is the linear part in δ? We have

$$g_t^*(\varepsilon,\delta) = \sum_{i=1}^n \delta_i\left(\sum_{A\subseteq\{1,\cdots,n\}, i\in A}(1-e^{-t(|A|-1)}\varepsilon_j)\right) + \Phi(\varepsilon,\delta)$$

where Φ is the remaining part. The way to write this is to write $\mathbb{E}\Phi(\varepsilon,\delta)\delta_i = 0$, for all $i = 1,\cdots,n$ since we know the extra part is orthogonal to all the linear parts, since the Walsh function determines an orthogonal basis.

So

$$\sum_{A\subseteq\{1,\cdots,n\}, i\in A}(1-e^{-t})e^{-t|A|}e^t\prod_{j\in A\setminus\{i\}}\hat{g}^*(A) = (e^t-1)\varepsilon_i e^{-t\Delta}\partial_i g^*$$

What does ∂_i do to g^*? It only keeps the is which belong to it, and it keeps it with the same coefficient. Then you hit it with $e^{-t\Delta}$, which corresponds to $e^{-t|A|}$. Then the last part is just the Walsh function.

All in all, you get that $g_t^*(\varepsilon,\delta) = (e^t-1)\sum_{i=1}^n \delta_i\varepsilon_i e^{-t\Delta}\partial_i g^*(\varepsilon) + \Phi(\varepsilon,\delta)$.

Now fix $\varepsilon\in\{\pm 1\}^n$. Recall the dual formulation of Rademacher type p: For every function $h^* : \{\pm 1\}^n \to X^*$, compute $\|(\sum_{i=1}^n \|\hat{h}^*(i)\|_X^{p^*})^{1/p^*}\|_{L_{r^*}(\{\pm 1\}^n, X^*)} \le \|h^*\|_{L_{r^*}(\{\pm 1\}^n, X^*)}$.

Then, applying this inequality to our situation with $g_t^*(\varepsilon,\delta)$ where we have fixed ε, we get

$$\left\|\left(\sum_{i=1}^n \|e^{-t\Delta}\partial_i g^*(\varepsilon)\|_{X^*}^{p^*}\right)^{1/p^*}\right\|_{L_{r^*}(\{\pm 1\}^n, X^*)} \lesssim (e^t-1)^{-1}\left(\mathbb{E}_\delta\|g_t^*(\varepsilon,\delta)\|_{X^*}^{r^*}\right)$$

Then, we get

$$\begin{aligned} \left\|\left(\sum_{i=1}^n \|e^{-t\Delta}\partial_i g^*\|_{X^*}^{p^*}\right)^{1/p^*}\right\|_{L_{r^*}(\{\pm 1\}^n, X^*)} &\lesssim (e^t-1)^{-1}\|g_t^*\|_{L_{r^*}(\{\pm 1\}^n\times\{\pm 1\}^n, X^*)} \\ &\lesssim (e^t-1)^{-1}\|g\|_{L_{r^*}(\{\pm 1\}^n, X^*)} \end{aligned}$$

Now we want to take $t \to \infty$. Let us look at $\|e^{-t\Delta}\partial_i g^*\|_{L_\infty(\{\pm 1\}^n, X^*)}$. $e^{-t\Delta}$ is a contraction, and so does ∂_i. We proved the first term is an averaging operator, which does not decrease norms. We also have that ∂_i is averaging a difference of two values, so it also does not decrease norm. Therefore, we can put a max on it and still get our inequality:

$$\max_{1\le i\le n} \|e^{-t\Delta}\partial_i g^*\|_{L_\infty(\{\pm 1\}^n, X^*)} \le \|g^*\|_{L_\infty(\{\pm 1\}^n, X^*)}$$

Then,

$$\left\| \left(\sum_{i=1}^n \|e^{-t\Delta}\partial_i g^*\|_{X^*}^a \right)^{1/a} \right\|_{L_b(\{\pm 1\}^n, X^*)} = \|(\|e^{-t\Delta}\partial_i g^*\|_{X^*})_{i=1}^n\|_{L_b(l_a^n(X^*))}$$

Define $S : L_{r^*}(X^*) \to L_{r^*}(l_{p^*}^n(X^*))$, and it maps

$$S(g^*) = (e^{-t\Delta}\partial_i g^*(\varepsilon))_{i=1}^n$$

The first inequality is the same thing as saying the operator norm of S

$$\|S\|_{L_{r^*}(X^*)\to L_{r^*}(l_{p^*}^n(X^*))} \lesssim \frac{1}{e^t-1}$$

We also know that

$$\|S\|_{L_\infty(X^*)\to L_\infty(l_\infty^n(X^*))} \le 1$$

We now want to interpolate these two statements (r^* and ∞) to arrive at q^*, as in the last remaining portion of the proof which is left.

Define $\theta \in [0,1]$ by $\frac{1}{q^*} = \frac{\theta}{p^*} + \frac{1-\theta}{\infty} = \frac{\theta}{p^*}$, so $\theta = \frac{p^*}{q^*}$.

By the vector-valued Riesz Interpolation Theorem (proof is identical to real-valued functions) (this is in textbooks),

$$\|S\|_{L_{a^*}(X^*)\to L_{a^*}(l_{q^*}^n(X^*))} \lesssim \frac{1}{(e^t-1)^\theta}$$

provided $\frac{1}{a^*} = \frac{\theta}{r^*} + \frac{1-\theta}{\infty}$, or $a^* = \frac{r^*}{\theta} = \frac{r^*q^*}{p^*}$. $\frac{q^*}{p^*} > 1$, as r ranges from $1 \to \infty$, so does r^*. In Pisier's paper, he says we can get any a^* that we want. However, this seems to be false. But this does not affect us: If we choose $r = p$, then we get $a = q$, which is all we needed to finish the proof.

So in 3.6.12, we take $r^* > \frac{q^*}{p^*}$, and this completes the claim.

Later I will either prove or give a reference for the interpolation portion of the argument.

Remember what we proved here and the definition of Enflo type p. You assign 2^n points in Banach space, to every vertex of the hyper cube. We deduce the volume of parallelpiped for any number of points. The open question is do we need to pass to this smaller value, and we needed it to get the integral to converge.

Chapter 4

Grothendieck's Inequality

1 Grothendieck's Inequality

Next week is student presentations, let's give some facts about Grothendieck's inequality. Let's prove the big Grothendieck theorem (this has books and books of consequences). Applications of Grothendieck's inequality is a great topic for a separate course.

The big Grothendieck Inequality is the following:

Theorem 4.1.1. *Big Grothendeick.*
There exists a universal constant K_G (the Grothendieck constant) such that the following holds. Let $A = (a_{ij}) \in M_{m\times n}(\mathbb{R})$, and $x_1, \dots, x_m, y_1, \dots, y_n$ be unit vectors in a Hilbert space H. Then there exist signs $\varepsilon_1, \dots, \varepsilon_n, \delta_1, \dots, \delta_n \in \{\pm 1\}^n$ such that if you look at the bilinear form

$$\sum_{i,j} a_{ij}\langle x_i, y_j\rangle \le K_G \sum_{i,j} a_{ij}\varepsilon_i\delta_j$$

So whenever you have a matrix, and two sets of high-dimensional vectors in a Hilbert space, there is at most a universal constant (less than 2 or 3) times the same thing but with signs.

Let's give a geometric interpretation. Consider the following convex bodies in $(\mathbb{R}^n)^2$. So A is the convex hull of all matrices of the form $conv(\{(\varepsilon_i\delta_j) : \varepsilon_1, \cdots, \varepsilon_m, \delta_1, \cdots, \delta_n \in \{\pm 1\}\}) \subseteq M_{m\times n}(\mathbb{R}) \cong \mathbb{R}^{mn}$. These matrices all have entries ± 1. This is a polytope.

$B = conv(\{\langle x_i, y_j\rangle : x_i, y_j$ unit vectors in a Hilbert space $\}) \subseteq M_{m\times n}(\mathbb{R})$. It's obvious that B contains A since A is a restriction to ± 1 entries.

Grothendieck says the latter half of the containment $A \subseteq B \subseteq K_G A$. If there is a point outside $K_G A$, you can find a separating hyperplane, which is a matrix. If B is not in $K_G A$, then there exists $\langle x_i, y_j\rangle \notin K_G A$ which means by separation there exists a matrix a_{ij} such that $\sum a_{ij}\langle x_i, y_j > K_G \sum a_{ij} c_{ij}$ for all $c_{ij} \in A$, which is a contradiction by Grothendieck's theorem taking $c_{ij} = \varepsilon_i\delta_j$.

We previously proved $l_\infty \to l_2$, the following is a harder theorem:

Lemma 4.1.2. *Every linear operator $T : l_\infty^n \to l_1^n$ satisfies for all $x_1, \cdots, x_m \in \ell_\infty^n$.*

$$\left(\sum_{i=1}^n \|Tx_i\|_{l_1^n}^2\right)^{1/2} \le K_G \cdot \|T\|_{l_\infty^n \to l_1^n} \max_{\|x\|_1 \le 1, x_i \in l_1^n} \left(\sum_{i=1}^n \langle x, x_i\rangle^2\right)^{1/2}$$

As an exercise, see how the lemma follows from Grothendieck, and how the little Grothendieck inequality follows from the big one in the case where a_{ij} is positive definite (so it's just a special case; there the best constant is $\sqrt{\pi/2}$).

Note that $\max_{\|x\|_1 \le 1}\langle x, x_i\rangle$ just gives $\|x_i\|_\infty$. But Grothendieck says for any number of $x_1, \cdots x_m$, we can for free improve our norm bound by a constant universal factor.

In the little Grothendieck inequality, what we did was prove the above fact for an operator for $l_1 \to l_2$, and we deduced Pietch domination from that conclusion: There is a probability measure on the unit ball of ℓ_1^n such that $\|Tx\|_1^2 \le K_G^2 \|T\|_{\ell_\infty^n \to \ell_1^n} \int_{B_{\ell_1^n}} \langle y^*, x\rangle^2 d\mu(y^*)$. If you know there exists such a probability measure, you know the above lemma, since you just do this for each x_i, and this is at most the maximum so you can just multiply.

This is an amazing fact though: Look at $T : \ell_\infty^n \to \ell_1^n$, and look at $L_2(\mu,)$. This is saying that if you think of an element in ℓ_∞ as a bounded function on the L_2 unit ball of its dual. You can think like this for any Banach space. Then, we have a diagram mapping from $\ell_2^n \to L_2(\mu)$ by identity into a Hilbert space subspace H of $L_2(\mu)$, and $S : H \to \ell_1^n$. You got these operators to factor through a Hilbert space H, so we form a commutative diagram and the norm of S is at most $\|S\| \le K_G \|T\|$.

This is how this is used. These theorems give you for free a Hilbert space out of nothing, and we use this a lot. After the presentations, from just this duality consequence, I'll prove one or two results in harmonic analysis and another result in geometry, and it really looks like magic. You can end up getting things like Parseval for free. We already saw the power of this in Restricted Invertibility.

Let's begin a proof of Grothendieck's inequality.

Proof. We will give a proof getting $K_G \le \frac{\pi}{2\log(1+\sqrt{2})} = 1.7\cdots$ (this is not from the original theorem). We don't actually know what K_G is. People don't really care what K_G is beyond the first digit. Grothendieck was interested in the worst configuration of points on the sphere, but it would not tell us what the configuration of points is. We want to know the actual number, or some description of the points. We know this is not the best bound since it doesn't give the worst configuration of points.

You do the following: The input was points $x_1, \cdots, x_m, y_1, \cdots, y_n \in$ unit sphere in Hilbert space H. We will construct a new Hilbert space H' and new unit vectors x_i', y_j' such that if z is uniform over the sphere of H' according to surface area measure, then the expectation of $\mathbb{E}\mathrm{sign}(\langle z, x_i'\rangle) * \mathrm{sign}(\langle z, y_j'\rangle) = \frac{2\log(1+\sqrt{2})}{\pi}\langle x_i, y_j\rangle$ for all i, j.

So we have a sphere with points x_i, y_j, and there is a way to nonlinearly transform into a sphere in the same dimension with x_i', y_j'. On the new sphere, if you take a uniformly random direction on the sphere, and look at a hyperplane defined by z, then the sign just indicates whether x_i', y_j' are on the same or opposite sides. Let's call ε_i the first "random sign", and δ_j the second "random sign". Then, $\mathbb{E}\sum a_{ij}\varepsilon_i\delta_j = \frac{2\log(1+\sqrt{2})}{\pi}\sum a_{ij}\langle x_i, y_j\rangle$. So there exists an instance (random construction) of ε_i, δ_j such that $\sum a_{ij}\langle x_i, y_j\rangle \le \frac{\pi}{2\log(1+\sqrt{2})}\sum a_{ij}\langle x_i, y_j\rangle$. If you succeed in making a bigger constant bound when bounding the expectation of the multiplication of the signs, you get a better Grothendieck constant. For the argument we will give, we will see that the exact number we get will be the best constant.

□

Undergrad Presentations.

2 Grothendieck's Inequality for Graphs (by Arka and Yuval)

We're going to talk about upper bounds for the Grothendieck constant for quadratic forms on graphs.

Theorem 4.2.1. *Grothendieck's Inequality.*

$$\sum_{i=1}^{n}\sum_{i=1}^{m} a_{ij}\langle x_i, y_i\rangle \le K_G \sum_{i=1}^{n}\sum_{i=1}^{m} a_{ij}\varepsilon_i\delta_j$$

where x_i, y_j are on the sphere. This is a bipartite graph where x_is form one side and y_js form the other side. Then you assign arbitrary weights a_{ij}. So you can consider arbitrary graphs on n matrices.

For every x_i, y_j, there exist signs ε_i, δ_j such that

$$\sum_{i,j\in\{1,\cdots,n\}} a_{ij}\langle x_i, y_j\rangle \le C\log n \sum a_{ij}\varepsilon_i\delta_j$$

We will prove an exact bound on the complete graph, where the Grothendieck constant is $C\log n$.

For a general graph, we will prove that $K(G) = \mathfrak{c}O(\log\Theta(G))$, where we will define the Θ function of a graph later.

Definition 4.2.2: $K(G)$ is the least constant K s.t. for all matrices $A : V\times V \to \mathbb{R}$,

$$\sup_{f:V\to S^{n-1}} \sum_{(u,v)\in E} A(u,v)\langle f(u), f(v)\rangle \le K \sup_{\varphi:V\to E} \sum_{(u,v)\in E} A(u,v)\varphi(u)\varphi(v)$$

Definition 4.2.3: The Gram representation constant of G.
Denote by $R(G)$ the infimum over constants R s.t. for every $f : V \to S^{n-1}$. There exists $F : V \to L_\infty^{[0,1]}$ such that for every $v \in V$ we have $\|F(v)\|_\infty \le R$ and $\langle f(u), f(v)\rangle = \langle F(u), F(v)\rangle = \int_0^1 F(u)(t)F(v)(t)dt$. For (u,v) which is an edge, you find the constant R so that you can embed functions f into F such that the ℓ_∞ norm is less than an explicit constant.

Lemma 4.2.4. *Let G be a loopless graph. Then $K(G) = R(G)^2$.*

Proof. Fix $R > R(G)$ and $f : V \to S^{n-1}$. Then there exists $F : V \to L_\infty[0,1]$ such that for every $v \in V$, we have $\|F(v)\|_\infty \le R$ and for $(u,v) \in E$ $\langle f(u), f(v)\rangle \le \langle F(u), F(v)\rangle$. Then,

$$\sum_{(u,v)\in E} A(u,v)\langle f(u), f(v)\rangle = \sum_{(u,v)\in E} A(u,v)\langle F(u), F(v)\rangle = \int \sum_{(u,v)\in E} A(u,v)F(u)(t)F(v)(t)dt$$
$$\le \int \sup_{g:V\to[-R,R]} \sum A(u,v)g(u)g(v)dt$$

We use definition and linearity to reverse the sum and integral. Now we use the loopless assumption to get to the definition of the right side of the Grothendieck inequality. We just need to fix the $[-R,R]$ to $[-1,1]$. We then get the last term above is equivalent to $R^2 \sup_{\varphi:V\to[-1,1]} \sum_{(u,v)\in E} A(u,v)\varphi(u)\varphi(v)$, which completes the proof.

Now, for each $f : V \to S^{n-1}$, consider $M(f) \subseteq \mathbb{R}^{|E|} = (\langle f(u), f(v)\rangle)_{(u,v)\in E}$. For each $\varphi : V \to \{-1,1\}$, define $M(\varphi) = (\varphi(u), \varphi(v))_{(u,v)} \in \mathbb{R}^{\varphi}$. Now we use the convex geometry interpretation of Grothendieck from the last class. Mimicking it, we write

$$\operatorname{conv}\{M(\varphi) : \varphi : V \to \{-1,1\}\} \subseteq \operatorname{conv}\left\{M(f) : f : V \to S^{n-1}\right\}$$

The implication of Grothendieck's inequality gives us

$$\operatorname{conv}\left\{M(f) : f : V \to S^{n-1}\right\} \subseteq K(G) \cdot \operatorname{conv}\{M(\varphi) : \varphi : V \to \{-1,1\}\}$$

So there exist weights $\{\lambda_g : g : V \to \{-1,1\}\}$ which satisfy $\sum_{q:V\to\{-1,1\}} \lambda_g = 1$, $\lambda_g \ge 0$ and for all $(u,v) \in E$ $\langle f(u), f(v)\rangle = \sum_{q:v\to\{-1,1\}} \lambda_g g(u)g(v)K(G)$.

Now consider

$$F(u) = \sqrt{K(g)} \cdot g(u)[\lambda_1 + \cdots + \lambda_{g-1}, \lambda_1 + \cdots + \lambda_g]$$

Then,

$$\langle F(u), F(v)\rangle = \sum_{g:\{V\to\{-1,1\}\}} K(G)g(u)g(v)\lambda_g$$

Then we consider our interval becomes $[-\sqrt{K(G)}, \sqrt{K(G)}]$, and thus $R(G) \le \sqrt{K(G)}$ and $R(G)^2 \le K(G)$. This proof can be modified for graphs with loops, but this is not quite true. □

An obvious corollary is that if H is a subgraph of G, then $R(H) \le R(G)$, and $K(H) \le K(G)$. This inequality is not obvious from the Grothendieck inequality directly, but is obvious going with our point of view.

Lemma 4.2.5. *Let K_n° denote the complete graph on n-vertices with loops. Then $R(K_n^\circ) = \mathfrak{c}O(\sqrt{\log n})$.*

Proof. Let σ be the normalized surface measure on S^{n-1}. By computation, there exists c s.t. $\sigma\left(\left\{x \in S^{n-1} : \|x\|_\infty \le c\sqrt{\frac{\log n}{n}}\right\}\right) \ge 1 - \frac{1}{2n}$, which can be calculated through integration.

When you get a function f on the sphere to the $n-1$ dimensional sphere, you want to find a rotation so that all these vectors have low coordinate vector value. We basically use the union bound. For each of the n vectors, the probability you get a vector with low valued last coordinate is $1 - 1/(2n)$, do this for all the vectors and you get probability greater than $1/2$. Then you can magnify this to get almost surely.

For every $x \in S^{n-1}$, the random variable on the orthogonal group $O(n)$ given by $U \to Ux$ is uniformly distributed on S^{n-1}. Thus for every $f : V \to S^{n-1}$, there is a rotation $U \in O(n)$ such that $\forall v \in V$ $\|U(f(v))\|_\infty \le c\sqrt{\frac{\log n}{n}}$.

We want $F(v)$ to be equal to the j^{th} coordinate of $Uf(v)$ on interval of length $1/n$. I.e., Let $F(u)(t) = (U(f(v)))\sqrt{n}$ on $\frac{j-1}{n} \le t\frac{j}{n}$.

Thus $R \le c\sqrt{\log n}$. □

Now I will prove the thing mentioned at the beginning:

Theorem 4.2.6. *$K(G) \le \log \chi(G)$, where $\chi(G)$ is the chromatic number.*

This theorem generalizes since on bipartite graphs $\chi(G)$ is a constant.

One thing we want to observe about the Grothendieck inequality is that it only cares about the Hilbert space structure. So we will just prove this in one specific (nice) Hilbert space and then we'll be done. Fix a probability space (Ω, P) such that $g_1, \cdots, g_n$ be i.i.d. standard Gaussians on Ω (for instance, Ω is the infinite product of $\mathbb{R}$).

Now define the Gaussian Hilbert space $H = \{\sum_{i=1}^\infty a_i g_i : \sum a_i^2 < \infty\} \subseteq L^2(\Omega)$. Every function in here will have mean zero since these are linear combinations of Gaussians. Note that the unit ball $B(H)$ consists of all Gaussian distributions with mean zero and variance at most 1 (since the variance is norm squared).

Let Γ be the left hand side of the Grothendieck inequality: $\sup_{f:V\to\{-1,1\}} \sum_{(u,v)\in E} A(u,v)\langle f(u), f(v)\rangle$, and let $\Delta = \sup_{f:V\to\{-1,1\}} \sum_{(u,v)\in E} A(u,v)\varphi(u)\varphi(v)$.

Definition 4.2.7: Truncation.
For all $M > 0$, $\psi \in L_2(\Omega)$, define the truncation of ψ at M to be

$$\psi^M(x) = \begin{cases} \psi(x) & |\psi(x)| \le M \\ M & \psi(x) \ge M \\ -M & \psi(x) \le -M \end{cases}$$

We're just cutting things off at the interval $[-M, M]$. So fix f maximizing Γ.

Lemma 4.2.8. *There exists Hilbert space H, $h : V \to H$, $M > 0$ with $\|h(v)\|_H^2 \le 1/2$ for all vertices $v \in V$, $M \lesssim \sqrt{\log \chi(G)}$, and we can now write Γ as a sum over all edges $\Gamma = \sum_{(u,v)} A(u,v)\langle f(u)^M, f(v)^M\rangle + \sum_{(u,v)\in E} A(u,v)\langle h(u), h(v)\rangle$, where the second term is an error term.*

Proof. Let $k = \chi(G)$. A fact for all $s : V \to l_2$ s.t. $\langle s(u), s(v)\rangle = \frac{-1}{k-1}$ for all $(u, v) \in E$, and $\|s(u)\| = 1$.

Define another Hilbert space $U = l_2 \oplus \mathbb{R}$. Define $t, \hat{t}$, two functions from $V \to U$. They are defined as follows:

$$t(u) = (\sqrt{\frac{k-1}{k}} s(u)) \oplus (\frac{1}{\sqrt{k}} e_1)$$

$$\hat{t}(u) = (-\sqrt{\frac{k-1}{k}} s(u)) \oplus (\frac{1}{\sqrt{k}} e_1)$$

Now, $t(u), \hat{t}(u)$ are unit vectors for all $(u, v) \in E$. Then $\langle t(u), t(v)\rangle = \langle \hat{(t)}(u), \hat{t}(v)\rangle = 0$. We also have $\langle t(u), \hat{t}(v)\rangle = \frac{2}{k}$.

Our goal is to prove that such a function exists. Set $H' = U \otimes L_2(\Omega)$. We now write down a function

$$h(u) = \frac{1}{4} t(u) \otimes (f(u) + f(u)^M) + k\hat{t}(u) \otimes (f(u) - f(u)^M)$$

It turns out that this function does everything we want. A couple of key facts: It's easy to check that the definition of Γ holds, defined in terms of $h(u)$ (for the error term). Checking $\|h(v)\|_H^2 \le 1/2$ holds is also possible. You can bound $\|h(u)\|^2 \le (1/2 + k\|f(u) - f(u)^M\|)^2$ after evaluating inner products. Now we use th eproperty of the Hilbert space. $f(u)$ is in the ball of H, with mean zero and variance at most 1. Then, $\|f(v) - f(v)^M\|^2$ is the probability that the Gaussian falls outside the interval $[-M, M]$. Therefore, from basic probability theory,

$$\|f(v) - f(v)^M\|^2 \le \frac{\sqrt{2}\pi}{\int} {}_M^\infty x^2 e^{-x^2/2} dx \le 2Me^{-M^2/2}$$

where the last part follows by calculus. This is great since if we pick $M \sim \sqrt{\log k}$. So this will decay super quickly. Picking $M = 8\sqrt{\log k}$ gives that the entire thing will be $\le 1/2$. That proves the lemma, so we're done. $\square$

This lemma is actually all we need. Suppose we have proved this. Now we can say the following. We can bound the first term by taking expectations: $\sum_{(u,v)\in E} A(u, v)\langle f(u)^M, f(v)^M\rangle = \mathbb{E}\sum_{(u,v)\in E} A(u, v) f(u)^M f(v)^M \le M^2\Delta$ by the definition of Δ.

For the second term, we write $\sum_{(u,v)\in E} A(u, v)\langle h(u), h(v)\rangle \le (\max_{v\in V} \|h(v)\|^2)\Gamma \le \frac{1}{2}\Gamma$. by re-scaling. You need to pull out each of these maxima separately, using linearity each time. You can also multiply each thing by $\sqrt{2}$ makes it be in $B(H)$. You just have to multiply by $\sqrt{2}$ and divide by $\sqrt{2}$ and they're not dependent anymore. By the conditions of the lemma, our h has norm at most $1/2$, so the This entire thing is Hilbert space independent, so the whole thing is $< \frac{\Gamma}{2}$. Thus $\Gamma \le M^2\Delta + \frac{1}{2}\Gamma \implies \Gamma \le 2M^2\Delta$.

This implies $\Gamma \lesssim \log \chi(G)\Delta$, which is exactly what we wanted to say, since the left hand side of the Grothendieck inequality is the first term, and the right hand side of the Grothendieck inequality is the second term with Grothendieck constant $\log \chi(G)$.

3 Noncommutative Version of Grothendieck's Inequality - Thomas and Fred

First we'll phrase the classical Grothendieck inequality in terms of an optimization problem. Given $A \in M_n(\mathbb{R})$, we can consider

$$\max_{\varepsilon_i,\delta_j\in} \sum_{i,j=1}^{n} A_{ij}\varepsilon_i\delta_j$$

In general this is hard to solve, so we do a semidefinite relaxation

$$\sup_{d \text{ dimensions}} \sup_{x,y\in(S^{d-1})^n} \sum_{i,j=1}^{n} A_{ij}\langle x_i, y_j\rangle$$

which implies that it's polynomially solveable, and Grothendieck inequality ensures you're only a constant factor off from the best.

We can give a generalization to tensors. Given $M \in M_n(M_n(\mathbb{R}))$ consider

$$\sup_{u,v\in O_n} \sum_{i,j,k,l=1}^{n} M_{ijkl}U_{ij}V_{kl}$$

Set $M_{iijj} = A_{ij}$, then we obtain

$$\sup \sum A_{ij}U_{ii}V_{jj} = \sup_{x,y\in\{-1,1\}^n} \sum_{i,j=1}^{n} A_{ij}x_iy_j = \max_{\varepsilon,\delta\in\{-1,1\}^n} \sum A_{ij}\varepsilon_i\delta_j$$

We relax to SDP over $\mathbb{R}$ of M:

$$\sup_{d\in\mathbb{N}} \sup_{X,Y\in O_n(\mathbb{R}^d)} \sum_{i,j,k,l=1}^{n} M_{ijkl}\langle X_{ij}, Y_{kl}\rangle$$

Recall that $U \in O_n$ means that

$$\sum_{k=1}^{n} U_{ik}U_{jk} = \sum_{k=1}^{n} U_{ki}U_{kj} = \delta_{ij}$$

If $X \in M_n(\mathbb{R}^d)$, let $XX^*, X^*X \in M_n(\mathbb{R})$ defined by

$$(XX^*)_{ij} = \sum_{k=1}^{n} \langle X_{ik}, Y_{jk}\rangle$$

$$(X^*X)_{ij} = \sum_{k=1}^{n} \langle X_{ki}, Y_{kj}\rangle$$

Then $O_n(\mathbb{R}^d) = \left\{X : M_n(\mathbb{R}^d) : XX^* = X^*X = I\right\}$.

We want to say something about when we relax the search space, we get within a constant factor of the non-relaxed version of the program. We will prove this in the complex case.

We will write

$$\mathrm{Opt}_{\mathbb{C}}(M) = \sup_{U,V\in U_n} \left| \sum_{i,j,k,l=1}^{n} M_{ijkl} U_{ij} \overline{V}_{kl} \right|$$

The fact that Opt is less than the SDP, Pisier proved thirty years after Grothendieck conjectured it. What we will actually prove is that the SDP solution is at most twice the optimal:

$$\mathrm{SDP}_{\mathbb{C}}(M) \leq 2\mathrm{Opt}_{\mathbb{C}}(M)$$

Theorem 4.3.1. *Fix $n, d \in \mathbb{N}$ and $\varepsilon \in (0,1)$. Suppose we have $M \in M_n(M_n(\mathbb{C}))$ and $X, Y \in U_n(\mathbb{C}^d)$ such that*

$$\left| \sum_{i,j,k,l=1}^{n} M_{ijkl} \langle X_{ij}, Y_{kl} \rangle \right| \geq (1-\varepsilon)\mathrm{SDP}_{\mathbb{C}}(M)$$

So what we're saying is suppose some SDP algorithm gave a solution satisfying this from input $X, Y \in U_n(\mathbb{C}^d)$. Then we can give a rounding algorithm which will output $A, B \in U_n$ such that

$$\mathbb{E}\left| \sum_{i,j,k,l=1}^{n} M_{ijkl} A_{ij} \overline{B}_{kl} \right| \geq (1/2 - \varepsilon)\mathrm{SDP}_{\mathbb{C}}(M)$$

Now what's the algorithm? It's a slightly clever version of projection. We have $X, Y \in U_n(\mathbb{C})$, and we want to get to actual unitary matrices. First sample $z \in \{1, -1, i, -i\}^d$ (complex unit cube). Then $x_z = \frac{1}{\sqrt{2}}\langle X, z\rangle$ (take inner products columnwise), similarly $Y_z = \frac{1}{\sqrt{2}}\langle Y, z\rangle$. Now we take the Polar Decomposition $A = U\Sigma V^*$ where U, V are unitary. The polar decomposition is just $A = (UV^*)(V\Sigma V^*)$, and then the first guy is unitary, and the second guy is PSD (think of it as $e^{i\theta} * r$).

Then we have $(A, B) = (U_z|X_z|^{it}, V_z|Y_z|^{-it})$ where t is sampled from the hyperbolic secant distribution. It's very similar to a normal distribution. The PDF is more precisely

$$\varphi(t) = \frac{1}{2}\mathrm{sech}(\frac{\pi}{2}t) = \frac{1}{e^{\pi t/2} + e^{-\pi t/2}}$$

When you have a positive definite matrix, you can raise it to imaginary powers, so we're rotating only the positive semidefinite part, and keeping the rotation side (UV^*). Note that the (A, B) are unitary. (You can also think of this as raising diagonals to it, these necessarily have eigenvalue of magnitude 1).

4-20-16

For $X, Y \in U_n(\mathbb{C}^d)$, do the following.

1. Sample $z \in \{\pm 1, \pm i\}$ uniformly. Sample t from the hyperbolic secant distribution.

2. Let $X_z = \frac{1}{\sqrt{2}}\langle x, z\rangle$, $Y_Z = \frac{1}{\sqrt{2}}\langle Y, z\rangle$.

3. Let $(A,B) := (U_Z|X_Z|^{it}, V_Z|Y_Z|^{-it}) \in U_n \times U_n$ where $X_Z = U_Z|X_Z|, Y_Z = U_Z|Y_Z|$ (polar decomposition).

$M(X,Y) = \sum_{i,j,k,l=1}^n M_{ijkl}\langle X_{ij}, Y_{kl}\rangle$. We want to show

$$\mathbb{E}_t[M(A,B)] \ge \left(\frac{1}{2}-\varepsilon\right) M(X,Y) \ge (\frac{1}{2}-\varepsilon)\mathrm{SDP}_{\mathbb{C}}(M)$$

We want to show that the rounded solution still has large norm.

It's possible that $|X_Z|$ has 0 eigenvalues. One solution is to add some Gaussian noise to the original X,Y because the set of non-invertible matrices is an algebraic set of measure 0. Alternatively, replace the eigenvalues by $\varepsilon \to 0$.

We have

$$\mathbb{E}_z[M(X_z,Y_z)] = \frac{1}{2}\mathbb{E}_z\left[\sum_{r,s=1}^d \overline{z_r} z_s \sum_{i,j,k,l=1}^n (X_{ij})r\overline{(Y_{kl})}_s\right] = \frac{1}{2}M(X,Y).$$

Now we analyze step 3. The characteristic function of the hyperbolic secant distribution is, for $a>0$,

$$\mathbb{E}[a^{it}] = \int a^{it}e(t)\,dt = \frac{2a}{Ha^2}$$

by doing a contour integral. Then

$$\mathbb{E}_t[a^{it}] = 2a - \mathbb{E}_t[a^{2+it}] \tag{4.1}$$
$$\mathbb{E}_t[A^{it}] = 2A - \mathbb{E}_t[A^{2+it}] \tag{4.2}$$
$$\begin{aligned}\mathbb{E}_t[A\otimes \overline{B}] &= \mathbb{E}_t[(U_z|X_z|^{it})\otimes(\overline{V_z}|Y_z|^{it})] &(4.3)\\ &= (U_z\otimes\overline{V_z})\,\mathbb{E}_t[(|X_z|\otimes|Y_t|)^{it}] &(4.4)\\ &= 2X_z\otimes\overline{Y_z} - \mathbb{E}_t[(U_z|X_z|^{2+it})\otimes(\overline{V_z|Y_z|}^{2-it})] &(4.5)\end{aligned}$$

We apply M.

$$\mathbb{E}_{z,t}[M(A,B)] = M(X,Y) - \mathbb{E}_{z,t}[M(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it})]$$

Because $A,B\in M_n(\mathbb{C})$, we can write $M(A,B)$ in terms of the tensor product,

$$\begin{aligned}M(A,B) &= \sum_{i,j,k,l=1}^n A_{ij}\overline{B_{kl}} &(4.6)\\ &= \sum_{i,j,k,l=1}^n M_{ijkl}(A\otimes\overline{B})_{(ij),(kl)} &(4.7)\end{aligned}$$

Claim 4.3.2 (Key claim)**.** *For all $t\in\mathbb{R}$,*

$$\left|\mathbb{E}_z\left[M(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it})\right]\right| \le \frac{1}{2}SDP_{\mathbb{C}}(M).$$

Proof. We have $F(t), G(t) \in M_n(\mathbb{C}^{\{\pm 1, \pm i\}^d})$. where

$$(F(t)_{jk})_z = \frac{1}{2^d}(U_z|X_z|^{2+it})_{jk} \tag{4.8}$$

$$(G(t)_{jk})_z = \frac{1}{2^d}(V_z|Y_z|^{2-it})_{jk}.M(F(t),G(t)) = \frac{1}{4^d}\sum_{z\in\{\pm1,\pm i\}^d} M(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it}) \tag{4.9}$$

$$= \mathop{\mathbb{E}}_z[M(U_z|X_z|^{2+it}, V_z|Y_z|^{2-it})]. \tag{4.10}$$

Lemma 4.3.3. *$F(t), G(t)$ satisfy*

$$\max\{\|F(t)F(t)^*\|, \|F(t)^*F(t)\|, \|G(t)G(t)^*\|, \|G(t)^*G(t)\|\}$$

Lemma 4.3.4. *Suppose $X, Y \in \mathcal{M}_n(\mathbb{C}^d)$ and $\max\{\|XX^*\|, \|X^*X\|, \|YY^*\|, \|Y^*Y\|\}$. Then there exist $R, S \in U_n(\mathbb{C}^{d+2n^2})$ such that $M(R,S) = M(X,Y) \le 1$ for every $M \in M_n(M_n(\mathbb{C}))$.*

From the two lemmas, there exist $R(t), S(t) \in U_n(\mathbb{C}^{d+2n^2})$ such that $M(R(t), S(t)) = M(\sqrt{2}F(t), \sqrt{2}G(t))$. Then

$$|M(F(t),G(t))| = \frac{1}{2}|M(R(t),S(t))| \le \frac{1}{2}2\mathrm{SDP}_{\mathbb{C}}(M) \tag{4.11}$$

$$F(t)F(t)^* = \frac{1}{4^d}\sum_{z\in\{\pm1,\pm i\}^d} U_z|X_z|^*U_z^* \tag{4.12}$$

$$= \mathop{\mathbb{E}}_z[U_z|X_z|^*U_2^*] = \mathop{\mathbb{E}}_z[(X_zX_z^*)]. \tag{4.13}$$

Claim 4.3.5. *For $W \in M_n(\mathbb{C}^d)$, define for each $v \in [d]$, $(W_r)_{ij} = (W_{ij})_r$, $W_z = \langle W, z\rangle$. Then*

$$\mathop{\mathbb{E}}_z[(W_zW_z^*)^2] = (WW^*)^2 + \sum_{r=1}^d W_r(W^*W - W_r^*W_r)W_r^*.$$

Note $W_z = \sum_{r=1}^d \Sigma_r W_r$ and

$$WW^* = \sum_{r=1}^d W_rW_r^* \qquad W^*W = \sum_{r=1}^d W_r^*W_r. \tag{4.14}$$

We compute

$$\mathop{\mathbb{E}}_z[(W_zW_z^*)^2] = \mathop{\mathbb{E}}_z\left[\sum_{=1}^{p,q,r,s} d\overline{z_p}z_q\overline{z_r}z_sW_pW_q^*W_rW_s^*\right] \tag{4.15}$$

$$= \sum_{p=1}^d W_pW_p^*W_pW_p^* + \sum_{\substack{p,q=1\\ p\ne q}}^n (W_pW_q^*W_qW_q^* + W_pW_q^*W_qW_p^*) \tag{4.16}$$

$$\begin{aligned}\sum_{p,q=1}^{d} W_pW_p^*W_qW_q^* + \sum_{p,q=1}^{d} W_pW_q^*W_qW_p^* - \sum_{p=1}^{d} W_pW_p^*W_pW_p^* \\ = \left(\sum_{p=1}^{d} W_pW_p^*\right)^2 + \sum_{p=1}^{d} W_p\left(\sum_{q=1}^{d} W_q^*W_q\right)W_p^* - \sum_{p=1}^{d} W_pW_p^*W_pW_p^*. \quad (4.17)\end{aligned}$$

Apply the claim with $W = \frac{1}{\sqrt{2}}X$. Recall that $XX^* = X^*X = I$. Then

$$F(t)F(t)^* + \frac{1}{4}\sum_{r=1}^{d} X_rX_r^*X_rX_r^* = \frac{1}{2}I = F(t)F(t)^* + \frac{1}{4}\sum_{r=1}^{d} X_r^*X_rX_r^*X_r$$

and similarly for G.

Proof of Lemma 2. Let $A = I - XX^*$, $B = I - X^*X$. Note $A, B \succeq 0$, $\mathrm{Tr}(A) = \mathrm{Tr}(\overline{B})$. We have

$$A = \sum_{i=1}^{n} \lambda_i(v_iv_i^*) \quad (4.18)$$

$$B = \sum_{j=1}^{n} \mu_j(v_jv_j^*) \quad (4.19)$$

$$\sigma = \sum_{i=1}^{n} \lambda_j = \sum_{j=1}^{n} \mu_j \quad (4.20)$$

$$R = X \oplus \left(\bigoplus_{i,j=1}^{n} \sqrt{\frac{\lambda_i\mu_j}{\sigma}}(u_iv_j^*)\right) \oplus O_{M_n(\mathbb{C}^{n^2})} \in M_n(\mathbb{C}^d \oplus \mathbb{C}^{n^2} \oplus \mathbb{C}^{n^2}) \quad (4.21)$$

$$S = Y \oplus O_{M_n(\mathbb{C}^{n^2})} \oplus \left(\bigoplus_{i,j=1}^{n} \sqrt{\frac{\lambda_i\mu_j}{\sigma}}(u_iv_j^*)\right) \quad (4.22)$$

Check $R \in U_n(\mathbb{C}^{d+2n^2})$,

$$RR^* = XX^* + A \quad (4.23)$$

$$R^*R = X^*X + B \quad (4.24)$$

$$M(R,S) = M(X,Y). \quad (4.25)$$

□

The σ disappears because $\sum \mu_j = \sigma$. □

The factor 2 in the noncommutative inequality is sharp. That the answer is 2 (rather than some strange constant) means there is something going on algebraically. The hyperbolic secant is the unique distribution that makes this work.

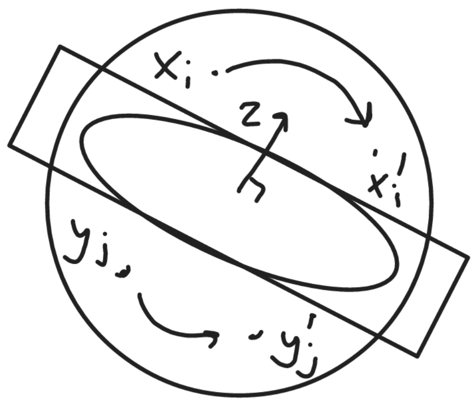
4-27

4 Improving the Grothendieck constant

Theorem 4.4.1 (Krivine (1977))**.**

$$K_G \le \frac{\pi}{2\ln(1+\sqrt{2})} \le 1.7...$$

The strategy is **preprocessing**.



There exist new vectors $x_1', \ldots, x_n', y_1', \ldots, y_n' \in \mathbb{S}^{2n-1}$ such that if $z \in \mathbb{S}^{2n-1}$ is chosen uniformly at random, then for all i, j,

$$\mathop{\mathbb{E}}_z \underbrace{(\operatorname{sign}\langle z, x_i'\rangle)}_{\varepsilon_i} \underbrace{(\operatorname{sign}\langle z, y_j'\rangle)}_{\delta_j} = \underbrace{\frac{2\ln(1+\sqrt{2})}{\pi}}_{c} \langle x_i, y_j\rangle .$$

Then

$$\sum a_{ij}\langle x_i, y_j\rangle = \mathbb{E}\left[\frac{\pi}{2\ln(1+\sqrt{2})}\sum a_{ij}\varepsilon_i\delta_j\right].$$

We will take vectors, transform them in this way, take a random point on the sphere, take a hyperplane orthogonal, and then see which side the points fall on.

Theorem 4.4.2 (Grothendieck's identity)**.** *For $x, y \in \mathbb{S}^k$ and z uniformly random on $\mathbb{S}^k$,*

$$\mathop{\mathbb{E}}_z[\operatorname{sign}(\langle x, z\rangle)\operatorname{sign}(\langle y, z\rangle)] = \frac{2}{\pi}\sin^{-1}(\langle x, y\rangle).$$

Proof. This is 2-D plane geometry.

The expression is 1 if both x, y are on the same side of the line, and -1 if the line cuts between the angle. $\square$

Once we have the idea to pre-process, the rest of the proof is natural.

Proof.

$$\mathop{\mathbb{E}}_z \left[\operatorname{sign}(\langle z, x_i' \rangle) \operatorname{sign}(\langle z, y_j' \rangle)\right] = \frac{2}{\pi} \sin^{-1}(\langle x_i', y_j' \rangle) = c \langle x_i, y_j \rangle \tag{4.26}$$

$$\langle x_i', y_j' \rangle = \sin\Big(\underbrace{\frac{\pi c}{2}}_{u} \langle x_i, y_j \rangle \Big). \tag{4.27}$$

We write the Taylor series expansion for sin.

$$\begin{aligned} \langle x_i', y_j' \rangle &= \sin(u \langle x_i, y_j \rangle) && (4.28)\\ &= \sum_{k=0}^{\infty} \frac{(u \langle x_i, y_j \rangle)^{2k+1}}{(2k+1)!} (-1)^k && (4.29)\\ &= \sum_{k=0}^{\infty} \frac{(-1)^k u^{2k+1}}{(2k+1)!} \left\langle x_i^{\otimes(2k+1)}, y_j^{\otimes(2k+1)} \right\rangle. && (4.30) \end{aligned}$$

(For $a, b \in \ell_2$, $a \otimes b = (a_i b_j)$, $a^{\otimes 2} = (a_i a_j)$, $b^{\otimes 2} = (b_i b_j)$, $\langle a^{\otimes 2}, b^{\otimes 2} \rangle = \sum a_i a_j b_i b_j = \langle a, b \rangle^2$.)

Define an infinite direct sum corresponding to these coordinates.

Define

$$x_i' = \bigoplus_{k=0}^{\infty} \left(\frac{(-1)^k u^{\frac{2k+1}{2}}}{\sqrt{(2k+1)!}} x_i^{\otimes 2k+1} \right) \tag{4.31}$$

$$y_j' = \bigoplus_{k=0}^{\infty} \left(\frac{u^{\frac{2k+1}{2}}}{\sqrt{(2k+1)!}} x_i^{\otimes 2k+1} \right) \in \bigoplus_{k=1}^{\infty} (\mathbb{R}^m)^{\otimes 2k+1} \tag{4.32}$$

$$\langle x_i', y_j' \rangle = \sum_{k=0}^{\infty} (-1)^k \frac{u^{2k+1}}{(2k+1)!} \left\langle x_i^{\otimes 2k+1}, y_j^{\otimes 2k+1} \right\rangle. \tag{4.33}$$

The infinite series expansion of sin generated an infinite space for us.

We check

$$\|x_i'\|^2 = \sum_{k=0}^{\infty} \frac{u^{(2k+1)}}{(2k+1)!} = \sinh(u) = \frac{e^u - e^{-u}}{2} = 1. \tag{4.34}$$

$\square$

We don't care about the construction, we care about the identity. All I want is to find x_i', y_j' with given $\langle x_i', y_j' \rangle$; this can be done by a SDP. We're using this infinite argument just to show existence.

A different way to see this is given by the following picture. Take a uniformly random line and look at the orthogonal projection of points on the line. Is it positive or negative. A priori we have vectors in $\mathbb{R}^n$ that don't have an order.

We want to choose orientations. A natural thing to do is to take a random projection onto $\mathbb{R}$ and use the order on $\mathbb{R}$. Krevine conjectured his bound was optimal.

What is so special about positive or negative? We can partition it into any 2 measurable sets. Any partition into 2 measurable sets produces a sign. It's a nontrivial fact that no partition beats this constant. This is a fact about measure theory.

The moment you choose a partition, it forced the construction of the x', y'. The whole idea is the partition; then the preprocessing is uniquely determined; it's how we reverse-engineered the theorem.

Here's another thing you can do. What's so special about a random line? The whole problem is about finding an orientation.

Consider a floor (plane) chosen randomly, look at shadows. If you partition the plane into 2 half-planes, you get back the same constant. We can generalize to arbitrary partitions of the plane. This is equivalent to an isoperimetric problem. In many such problems the best thing is a half-space. We can look at higher-dimensional projections. It seems unnatural to do this—except that you gain!

In $\mathbb{R}^2$ there is a more clever partition that beats the halfspace! Moreover, as you take the increase the dimension, eventually you will converge to the Grothendieck constant. The partitions look like fractal objects!

5 Application to Fourier series

This is a classical theorem about Fourier series. Helgason generalized it to a general abelian group.

Definition 4.5.1: Let $S = \mathbb{R}/\mathbb{Z}$. The space of continuous functions is $C(S')$. Given $m : \mathbb{Z} \to \mathbb{R}$ (the multiplier), define

$$\Lambda_m : C(\mathbb{S}') \to \ell_\infty \tag{4.35}$$

$$\Lambda_m(f) = (m(n)\widehat{f}(n))_{n\in\mathbb{Z}}. \tag{4.36}$$

For which multipliers m is $\Lambda_m(f) \in \ell_1$ for every $f \in C(\mathbb{S}^1)$?

A more general question is, what are the possible Fourier coefficients of a continuous functions? For example, if $\log n$ works, then $\sum |\widehat{f}(n)||m(n)| < \infty$. This is a classical topic.

An obvious sufficient condition is that $m \in \ell_2$, by Cauchy-Schwarz and Parseval.

$$\sum_{n\in\mathbb{Z}} |m(n)\widehat{f}(n)| \le \left(\sum m(n)^2\right)^{\frac{1}{2}} \underbrace{\left(\sum |\widehat{f}(n)|^2\right)^{\frac{1}{2}}}_{\|f\|_2 \le \|f\|_\infty} \tag{4.37}$$

$$\|\Lambda_m(f)\|_{\ell_1} \le \|m\|_2 \|f\|_\infty . \tag{4.38}$$

This theorem says the converse.

Theorem 4.5.2 (Orlicz-Paley-Sidon)**.** *If $\Lambda_m(f) \in \ell_1$ for all $f \in C(\mathbb{S}^1)$, then $m \in \ell_2$.*

We know the fine line of when Fourier coefficients of continuous functions converge.

We can make this theorem quantitative. Observe that if you know $\Lambda_m : C(\mathbb{S}^1) \to \ell_1$, then Λ_m has a closed graph (exercise). The closed graph theorem (a linear operator between Banach spaces with closed graph is bounded) says that $\|\Lambda_m\| < \infty$.

Corollary 4.5.3. $\sum |m(n)\widehat{f}(n)| \le K \|f\|_\infty$.

We show $\|\Lambda_m\| \le \|m\|_2 \le K_G \|\Lambda_m\|$.

We will use the following consequence of Grothendieck's inequality (proof omitted).

Corollary 4.5.4. *Let $T : \ell_\infty^n \to \ell_1^m$. For all $x_1, \ldots, x_n \in \ell_\infty$,*

$$\left(\sum_i \|Tx_i\|_1^2 \right)^{\frac{1}{2}} \le K_G \|T\| \sup_{y \in \ell_1} \sum_{i=1}^n \langle y, x_i \rangle^2.$$

Equivalently, there exists a probability measure μ on $[n]$ such that $\|Tx\| \le K_G \|T\| \int K_G \|T\| \left(\int x_j^2 \, d\mu(j) \right)$

Proof. Use duality. To get the equivalence, use the same proof as in Piesch Domination. $\square$

Proof. Given $\Lambda_m : C(\mathbb{S}^1) \to \ell_1$, there exists μ on $\mathbb{S}^1$ such that for every f, letting $f_\theta(x) = f(e^{i\theta}x)$,

$$\left(\sum |m(n)\widehat{f}_\theta(n)| \right)^2 \le K_G^2 \|\Lambda_m\|^2 \int_{S^1} (f_\theta(x))^2 \, d\mu(x) \tag{4.39}$$

$$\left(\sum |m(n)\widehat{f}(n)| \right)^2 \le K_G^2 \|\Lambda_m\|^2 \int_{S^1} (f(x))^2 \, d\mu(x). \tag{4.40}$$

Apply this to the following trigonometric sum

$$f(x) = \sum_{n=-N}^{N} m(n) e^{-in\theta},$$

to get

$$\left(\sum_{n=-N}^{N} m(n)^2 \right)^2 \le K_G^2 \|\Lambda\|_m^2 \sum_{n=-N}^{N} m(n)^2.$$

$\square$

RIP also had this kind of magic.

6 ? presentation

Theorem 4.6.1. *Let (X, d) be a metric space with $X = A \cup B$ such that*

- *A embeds into ℓ_2^a with distortion D_A, and*
- *B embeds into ℓ_2^b with distortion D_B.*

Then X embeds into ℓ_2^{a+b+1} with distortion at most $7D_AD_B + 5(D_A + D_B)$.

Furthermore, given $\psi : A \to \ell_2^a$ and $\varphi_B : B \to \ell_2^b$ with $\|\varphi_A\|_{Lip} \le D_A$ and $\|\varphi_B\|_{Lip} \le D_B$, there is an embedding $\Psi : X \hookrightarrow \ell_2^{a+b+1}$ with distortion $7D_AD_B + 5(D_A + D_B)$ such that $\|\Psi(u) - \Psi(v)\| \ge \|\varphi_A(u) - \varphi_A(v)\|$ for all $u, v \in A$.

For $a \in A$, let $R_a = d(a, B)$ and for $b \in B$, let $R_b = d(A, b)$.

Definition 4.6.2: For $\alpha > 0$, $A' \subseteq A$ is an α**-cover** for A with respect to B if

1. For every $a \in A$, there exists $a' \in A'$ where $R_{a'} \le R_a$ and $d(a, a') \le \alpha R_a$.
2. For distinct $a_1', a_2' \in A'$, $d(a_1', a_2') \ge \alpha \min(R_{a_1'}, R_{a_2'})$.

This is like a net, but the ε of the net is a function of the distance to the set. The requirement is weaker for points that are farther away.

Lemma 4.6.3. *For all $\alpha > 0$, there is an α-cover A' for A with respect to B.*

Proof. Induct. The base case is ϕ, which is clear.

Assume the lemma holds for $|A| < k$; we'll show it holds for $|A| = k$. Let $u \in A$ be the point closest to B. Let $Z = A \backslash B_{\alpha R_u}(u) =: B$; we have $|Z| < |A|$. Let Z' be a cover of Z, and let $A' = Z' \cup \{u\}$.

We show that A' is an α-cover. We need to show the two properties.

1. Divide into cases based on whether $a \in B$ or $a \notin B$.
2. For $a_1', a_2' \in A'$, if both are in z', we're done.

   Otherwise, without loss of generality, $a_1' = u$ and $a_2' \in Z'$.

□

Lemma 4.6.4. *Define $f : A' \to B$ to send every point in the α-cover to a closest point in B, $d(a', f(a')) = R_{a'}$.*

Then $\|f\|_{Lip} \le 2\left(1 + \frac{1}{\alpha}\right)$. By the triangle inequality,

$$d(f(a_1'), f(a_2')) \le d(f(a_1'), a_1') + d(a_1', a_2') + d(a_2', f(a_2')).$$

We have

$$\begin{aligned} Ra_1' + Ra_2' + d(a_1', a_2') &= 2\min(R_{a_1'}, R_{a_2'}) + |R_{a_1'} + R_{a_2'} + d(a_1', a_2') && (4.41)\\ &\le \frac{1}{\alpha} d(a_1', a_2') + d(a_1', a_2') + d(a_1', a_2') && (4.42)\\ &= 2\left(1 - \frac{1}{\alpha}\right) d(\alpha_1', \alpha_2') && (4.43) \end{aligned}$$

Let $\varphi_B : B \hookrightarrow \ell_2^b$ be an embedding with $\|\varphi\|_{\text{Lip}} \le D_B$; it is noncontracting.

Lemma 4.6.5. *There is $\psi : X \to \ell_2^b$ such that*

1. *For all $a_1, a_2 \in A$,*
$$\|\psi(a_1) - \psi(a_2)\| \le 2\left(1 + \frac{1}{\alpha}\right) D_A D_B d(\alpha_1, \alpha_2)$$
2. *For all $b_1, b_2 \in B$,*
$$d(b_1, b_2) \le \|\psi(b_1) - \psi(b_2)\| = \|\varphi_B(b_1) - \varphi_B(b_2)\| \le D_B d(b_1, b_2).$$
3. *For all $a \in A$, $b \in B$, $d(a, b) - (1+\alpha)(2D_A D_B + 1)R_a \le \|\psi(a) - \psi(b)\| \le 2(1 + \alpha)(D_A D_B + (2+\alpha)D_B)d(a, b)$.*

Proof. Let $g = \varphi_B f \varphi_A^{-1}$.

We have maps

$$\begin{tikzcd}
A \arrow[r, hook, "c"] \arrow[d, "\varphi_A"] & A' \arrow[r, "f"] \arrow[d, "\varphi_A"] & B \arrow[d, "\varphi_B"] & \\
\varphi(A) & \varphi(A') \arrow[l, "c"'] \arrow[r] & \varphi(B) \arrow[r, hook] & \ell_2^b.
\end{tikzcd}$$

Now
$$\|g\|_{\text{Lip}} \le \|\varphi_B\|_{\text{Lip}} \|f\|_{\text{Lip}} \left\|\varphi_A^{-1}\right\|_{\text{Lip}} \le D_B 2\left(1 + \frac{1}{\alpha}\right)$$
By the Kirszbraun extension theorem 3.6.1, construct $\widetilde{g} : \ell_2^a \to \ell_2^b$ with
$$\|\widetilde{g}\|_{\text{Lip}} \le D_B 2\left(1 + \frac{1}{\alpha}\right).$$

Define $\psi(x) = \begin{cases} \widetilde{g}(\varphi_A(x)), & \text{if } x \in A \\ \varphi_B(x), & \text{if } x \in B. \end{cases}$.

We show the three parts.

1.
$$\begin{aligned}
\|\psi(a_1) - \psi(a_2)\| &= \|\widetilde{g}(\varphi_A(a_1)) - \widetilde{g}(\varphi_A(a_2))\| && (4.44)\\
&\le \|\widetilde{g}\|_{\text{Lip}} \|\varphi_A\|_{\text{Lip}} d(a_1, a_2). && (4.45)
\end{aligned}$$

2. This is clear.

3. Let $b = f(a')$. Then $d(a', b') \le R_a$, $\psi(a') = \psi(b')$. We have
$$\begin{aligned}
\|\psi(a) - \psi(b)\| &\le \|\psi(a) - \psi(a')\| - \|\psi(a') - \psi(b')\| - \|\psi(b) - \psi(b')\| && (4.46)\\
\|\psi(a) - \psi(b)\| &\le 2\left(1 + \frac{1}{\alpha}\right) D_A D_B d(a, a') + D_B d(b, b') && (4.47)\\
d(a, a') &\le \alpha R_a \le \alpha d(a, b) && (4.48)\\
d(b, b') &\le d(b, a) + d(a, a') + d(a', b') && (4.49)\\
&\le (2+\alpha) d(a, b). && (4.50)
\end{aligned}$$

This shows the first half of (3).

For the other inequality, use the triangle inequality against and get

$$\begin{aligned}\|\psi(a)-\psi(b)\| &\geq \|\psi(b)-\psi(b')\| - \|\psi(a')-\psi(b')\| - \|\psi(a')-\psi(a)\| && (4.51)\\ &\geq d(b,b') - 2(1+\alpha)D_A D_B R_a. && (4.52)\end{aligned}$$

□

Let

$$\begin{aligned}\psi_B &= \psi && (4.53)\\ \beta &= (1+\alpha)(2D_A D_B + 1) && (4.54)\\ \gamma &= \left(\frac{1}{2}\right)\beta && (4.55)\\ \psi_\Delta &: X \to \mathbb{R} && (4.56)\\ \psi_A(a) &= \gamma R_a, \qquad a \in A && (4.57)\\ \psi_A(b) &= -\gamma R_b, \qquad b \in B && (4.58)\\ \Psi &: X \to \ell^{a+b+1} && (4.59)\\ \Psi(x) &= \psi_A \oplus \psi_B \oplus \psi_\Delta \in \ell_2^{a+b+1}. && (4.60)\end{aligned}$$

For $a_1, a_2 \in A$,

$$\|\Psi(a_1)-\Psi(a_2)\| \geq \|\Psi_A(a_1)-\Psi_A(a_2)\| \geq d(a_1,a_2).$$

For $a \in A, b \in B$,

$$\|\Psi(a)-\Psi(b)\|^2 = \|\psi_A(a)-\psi_A(b)\|^2 + \|\psi_B(a)-\psi_B(b)\|^2 + \|\psi_A(a)-\psi_A(b)\|^2.$$

$$\begin{aligned}\|\psi_A(a)-\psi_A(b)\| &\geq d(a,b) - \beta R_b && (4.61)\\ \|\psi_B(a)-\psi_B(b)\| &\geq d(a,b) - \beta R_a && (4.62)\\ \|\psi_A(a)-\psi_A(b)\| &= \gamma(R_a + R_b). && (4.63)\end{aligned}$$

Claim 4.6.6. *We have $\|\psi(a)-\psi(b)\| \geq d(a,b)$.*

Proof. Without loss of generality $R_a \subseteq R_b$. Consider 3 cases.

1. $\beta R_b \leq d(a,b)$.
2. $\beta R_a \leq d(a,b) \leq \beta R_b$.
3. $d(a,b) \leq \beta R_a$.

Consider case 2. The other cases are similar.

$$\|\psi(a)-\psi(b)\|^2 \geq (d(a,b)-\beta R_a)^2+\beta^2(R_a+R_b)^2/2 \tag{4.64}$$
$$= d(a,b)^2-2\beta d(a,b)R_a+\frac{\beta^2}{2}(3R_a^2+2R_aR_b+R_b^2) \tag{4.65}$$
$$\geq d(a,b)^2-2\beta R_aR_b+\frac{\beta^2}{2}(3R_a^2+2R_aR_b+R_b^2) \tag{4.66}$$
$$= d(a,b)^2+\beta^2((\sqrt{3}R_a-R_b)^2+2(\sqrt{3}+R_aR_b))/2 \tag{4.67}$$
$$> d(a,b) \tag{4.68}$$

□

For $a_1, a_2 \in X$,

$$\|\Psi(a_1)-\Psi(a_2)\| = \|\psi_A(a_1)+\psi_A(a_2)\|^2+\|\psi_B(a_1)-\psi_B(a_2)\|^2+\|\psi_B(a_1)-\psi_A(a_2)\| \tag{4.69}$$
$$\leq \left(D_A^2+4\left(1+\frac{1}{\alpha}\right)^2 D_A^2D_B^2\right)d(a_1,a_2)^2+\gamma^2(R_{a_1}-R_{a_2}) \tag{4.70}$$
$$\leq \left(D_A^2+4\left(1+\frac{1}{\alpha}\right)^2 D_A^2D_B^2+\gamma^2\right)d(a_1,a_2)^2 \tag{4.71}$$
$$|R_a-R_{a_2}| \leq d(a_1,a_2) \tag{4.72}$$
$$\|\Psi(b_1)-\Psi(b_2)\|^2 \leq \left(D_B^2+4\left(1+\frac{1}{\alpha}\right)^2 D_A^2D_B^2+\gamma^2\right)d(a,b)^2 \tag{4.73}$$
$$|\Psi(a)-\Psi(b)|^2 \leq \|\psi_A(a)-\psi_B(b)\|^2+\|\psi_B(a)-\psi_B(b)\|^2+\|\psi_A(a)-\psi(b)\|^2 \tag{4.74}$$
$$\leq \cdots \gamma^2(R_a+R_b)^2 \leq \cdots \tag{4.75}$$

$R_a, R_b \leq d(a,b)$.

Chapter 5

Lipschitz extension

'

1 Introduction and Problem Setup

We give a presentation of the paper "On Lipschitz Extension From Finite Subsets", by Assaf Naor and Yuval Rabani, (2015). For the convenience of the reader referencing the original paper, we have kept the numbering of the lemmas the same.

Consider the setup where we have a metric space (X, d_X) and a Banach space $(Z, \|\cdot\|_Z)$. For a subset $S \subseteq X$, consider a 1-Lipschitz function $f : S \to Z$. Our goal is to extend f to $F : X \to Z$ without experiencing too much growth in the Lipschitz constant $\|F\|_{Lip}$ over $\|f\|_{Lip}$.

Definition 5.1.1: $e(X, S, Z)$ and its variants.
Define $e(X, S, Z)$ to be the infimum over the sequence of K satisfying $\|F\|_{Lip} \leq K\|f\|_{Lip}$ (i.e., $e(X, S, Z)$ is the least upper bound for $\frac{\|F\|_{Lip}}{\|f\|_{Lip}}$ for a particular S, X, Z).

Then, define $e(X, Z)$ to be the supremum over all subsets S for $e(X, S, Z)$: So of all subsets, what's the largest least upper bound for the ratio of Lipschitz constants?

We may also want to consider supremums over $e(X, S, Z)$ for S with a fixed size. We can formulate this in two ways. $e_n(X, Z)$ is the supremum of $e(X, S, Z)$ over all S such that $|S| = n$. We can also describe $e_\epsilon(X, Z)$ as the supremum of $e(X, S, Z)$ over all S which are ϵ-discrete in the sense that $d_X(x, y) \geq \epsilon \cdot {}_{\text{diam}}(S)$ for distinct $x, y \in S$ and some $\epsilon \in [0, 1]$.

Definition 5.1.2: Absolute extendability.
We define $\mathbf{ae}(n)$ to be the supremum of $e_n(X, Z)$ over all possible metric spaces (X, d_X) and Banach spaces $(Z, \|\cdot\|_Z)$. Identically, $\mathbf{ae}(\epsilon)$ is the supremum of $e_\epsilon(X, Z)$ over all (X, d_X) and $(Z, \|\cdot\|_Z)$.

From now on, we will primarily discuss subsets $S \subseteq X$ with size $|S| = n$. Bounding the supremum, the absolute extendability $\mathbf{ae}(n) < K$ allows us to make general claims about the extendability of maps from metric spaces into Banach spaces. **Any** Banach-space valued 1-Lipschitz function defined on metric space (M, d_M) can therefore be extended to **any** metric space M' such that M' contains M (up to isometry; as long as you can embed M in M' with an injective distance preserving map) such that the Lipschitz constant of the extension is at most K.

Therefore, for the last thirty years, it has been of interest to understand upper and lower bounds on $\mathbf{ae}(n)$, as we want to understand the asymptotic behavior as $n \to \infty$. In the 1980s, the following upper and lower bound were given by Johnson and Lindenstrauss and Schechtman:

$$\sqrt{\frac{\log n}{\log\log n}} \lesssim \mathbf{ae}(n) \lesssim \log n$$

In 2005, the upper bound was improved:

$$\sqrt{\frac{\log n}{\log\log n}} \lesssim \mathbf{ae}(n) \lesssim \frac{\log n}{\log\log n}$$

In this talk, we improve the lower bound for the first time since 1984 to

$$\sqrt{\log n} \lesssim \mathbf{ae}(n) \lesssim \frac{\log n}{\log\log n}$$

1.1 Why do we care?

This improvement is of interest primarily not because of the removal of a $\sqrt{\log\log n}$ term in the denominator. It is due to the fact that the approach taken to get the lower bound provided by Johnson-Lindenstrauss 1984 has an inherent limitation. The approach of Johnson-Lindenstrauss to get the lower bound is to prove the nonexistance of linear projections of small norm. By considering a specific case for f, X, S, Z, we can get a lower bound on $\mathbf{ae}(n)$. Consider a Banach space $(W, \|\cdot\|_W)$ and let $Y \subseteq W$ be a k-dimensional linear subspace of W with N_ϵ an ϵ-net in the unit sphere of Y, and then define $S_\epsilon = N_\epsilon \cup \{0\}$. Fix $\epsilon \in (0, 1/2)$. We take $f : S_\epsilon \to Y$ to be the identity mapping, and wish to find an extension to $F : W \to Y$. Then, in our setup, we let $X = W$, $S = S_\epsilon$, $Z = Y$. We seek to bound the magnitude of the Lipschitz constant of F, call it L. Johnson-Lindenstrauss prove that for $\epsilon \lesssim \frac{1}{k^2}$, there **exists** a linear projection $P : W \to Y$ with $\|P\| \lesssim L$. We can now proceed to lower bound L by lower-bounding $\|P\|$ for all P. The classical Kadec'-Snobar theorem says that there always exists a projection with $\|P\| \le \sqrt{k}$. Therefore, the best (largest) possible lower bound we could get will be $L \gtrsim \sqrt{k}$ by Kadec'-Snobar. But this is bad:

Taking $n = |S_\epsilon|$, by bounds on ϵ-nets we get $k \asymp \frac{\log n}{\log(1/\epsilon)}$ which implies

$$L \gtrsim \sqrt{\frac{\log n}{\log(1/\epsilon)}}$$

In order to get the lower bound on $\mathbf{ae}(n)$ of $\sqrt{\log n}$, we must take ϵ to be a universal constant. However, from a lemma by Benyamini (in our current setting), $L \lesssim e_\epsilon(X, Z) \lesssim 1/\epsilon = O(1)$, which means that any lower bound we get on L will be too small (and won't even tend to ∞). Therefore, we must make use of nonlinear theory to get the $\sqrt{n}$ lower bound on $\mathbf{ae}(n)$.

1.2 Outline of the Approach

Let us formally state the theorem, and then give the approach to the proof.

Theorem 5.1.3. *Theorem* 1.
For every $n \in \mathbb{N}$ we have $\boldsymbol{ae}(n) \gtrsim \sqrt{\log n}$.

We give a metric space X, a Banach space Z, a subset $S \subseteq X$, a function $f : S \to Z$ such that f extends to $F : X \to Z$ where $\|F\|_{Lip} \leq K\|f\|_{Lip}$.

Let V_G be the vertices of a finite graph G with distance metric the shortest path metric d_G where edges all have length 1.

We define our metric space $X = (V_G, d_{G_r(S)})$ where $G_r(S)$ is the r-magnification of the shortest path metric on V_G. S is an n-vertex subset $(S, d_{G_r(S)})$. Our Banach space $Z = (\mathbb{R}_0^X, \|\cdot\|_{W_1(X, d_{G_r(S)})})$ is equipped with the Wasserstein-1 norm induced by the r-magnification of the shortest path metric on the graph. Note that $\mathbb{R}_0^X$ is just weight distributions on the vertices of X which sum to zero in the image. Our $f : S \to \mathbb{R}_0^S \subseteq Z$, and we extend to $F : X \to Z$. We will show how to choose r and $|S|$ optimally to get the result.

The rest of my section of the talk will give the requisite definitions and lemmas to understand the full proof.

2 r-Magnification

Definition 5.2.1: r-magnification of a metric space.
Given metric space (X, d_X) and $r > 0$, for every subset $S \subseteq X$ we define $X_r(S)$ as a metric space on the points of X equipped with the following metric:

$$d_{X_r(S)}(x, y) = d_X(x, y) + r|\{x, y\} \cap S|$$

and where $d_{X_r(S)}(x, x) = 0$. All this is saying is that when we have distinct points $x, y \in S$, we have the metric is $2r + d_X(x, y)$, when one point is in S and one point is outside, we have $r + d_X(x, y)$, and when both x, y are outside, the metric is unchanged.

The significance of this definition is as follows: It's easier for functions on S to be Lipschitz (we enlarge the denominator) without affecting functions on $X \setminus S$. Thus, there are more potential f we can draw from which satisfy 1-Lipschitzness which can have potentially large Lipschitz extensions (i.e., large K) since we don't make it easier to be Lipschitz on $X \setminus S$ (which we must deal with in the extension space).

However, we can't make r too large: the minimum distance between x, y in S becomes close to diam(S) under r-magnification as r increases. Let us assume the minimum distance between x, y is 1 (as it would be in an undirected graph with an edge between x, y under the shortest path metric). Particularly, for distinct $x, y \in S$, since $\text{diam}(S, d_{X_r(S)}) = 2r + \text{diam}(S, d_X)$,

$$d_{X_r(S)}(x, y) \geq 2r + 1 = \frac{2r + 1}{2r + \text{diam}(S, d_X)} \cdot \text{diam}(S, d_{X_r(S)})$$

Then recall that $e_\epsilon(X,Z)$ is the supremum over S such that are ϵ-discrete, where here, $\epsilon = \frac{2r+1}{2r+\text{diam}(S,d_X)}$. Earlier we saw a bound that

$$e_\epsilon(X,Z) \lesssim 1/\epsilon = \frac{2r+\text{diam}(S,d_X)}{2r+1} \le 1 + \frac{\text{diam}(S,d_X)}{r}$$

Thus, if we make r too large, we again are bounding $e_\epsilon(X,Z) \lesssim 1 = O(1)$, which means our choice of X and Z is not good to get a large lower bound (again, we're not even going to ∞).

Thus we must balance our choice of r appropriately.

3 Wasserstein-1 norm

Now we come to the second part of our choice of Z. Note that we will define $\mathbb{R}_0^X$ to be the set of functions on the points of X such that for each $f \in \mathbb{R}_0^X$, $\sum_{x\in X} f(x) = 0$. We use e_x to denote the indicator weight map with 1 at point x and 0 everywhere else.

Definition 5.3.1: Wasserstein-1 Norm.
The Wasserstein-1 norm is the norm induced by the following origin-symmetric convex body in finite metric space (X, d_X):

$$K_{(X,d_X)} = \text{conv}\left\{\frac{e_x - e_y}{d_X(x,y)} : x, y \in X, x \neq y\right\}$$

This is a unit ball on $\mathbb{R}_0^X$. We denote the induced norm by $\|\cdot\|_{W_1(X,d_X)}$.

We can give an equivalent (proven with the Kantorovich-Rubinstein duality theorem) definition of the Wasserstein-1 distance:

Definition 5.3.2: Wasserstein-1 distance and norm.
Let $\Pi(\mu,\nu)$ be all measures on π on $X \times X$ such that

$$\sum_{y\in X} \pi(y,z) = \nu(z)$$

for all $z \in X$ and

$$\sum_{z\in X} \pi(y,z) = \mu(y)$$

for all $y \in X$. Then, the Wasserstein-1 distance (earthmover) is

$$W_1^{d_X}(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \sum_{x,y\in X} d_X(x,y)\pi(x,y)$$

In the case that $\mu = \nu$, we automatically have $(\mu\times\nu)/\mu(X) \in \Pi(\mu,\nu)$ (normalizing by one measure trivially gives the other), so Π is nonempty. The norm induced by this metric for $f \in \mathbb{R}_0^X$ is

$$\|f\|_{W_1(X,d_X)} = W_1^{d_X}(f^+, f^-)$$

where $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$ (since $\sum_{x\in X} f(x) = 0$, we need to make sure that μ, ν are both nonnegative measure). Futhermore, we have $\sum_{x\in X} f^+(x) = \sum_{x\in X} f^-(x)$ (same total mass), which means that $\Pi(f^+, f^-)$ is nonempty.

Definition 5.3.3: $\ell_1(X)$ norm.
Note that using standard notation, we can also define an ℓ_1 norm on f by

$$\|f\|_{\ell_1(X)} = \sum_{x\in X} |f(x)| = \sum_{x\in X} f^+(x) + f^-(x)$$

Thus, for our restrictions on f, $\sum_{x\in X} f^+(x) = \sum_{x\in X} f^-(x) = \|f\|_{\ell_1(X)}/2$.

Now we give a simple lemma which gives bounds for the Wasserstein-1 norm induced by the r-magnification of a metric on X.

Lemma 5.3.4. *Lemma 7.*
For (X, d_X) a finite metric space, we have

1. $\|e_x - e_y\|_{W_1(X,d_X)} = d_X(x, y)$ *for every* $x, y \in X$.
2. *For all* $f \in \mathbb{R}^X_0$,
$$\frac{1}{2} \min_{x,y\in X; x\neq y} d_X(x, y)\|f\|_{\ell_1(X)} \leq \|f\|_{W_1(X,d_X)} \leq \frac{1}{2}\,diam(X, d_X)\|f\|_{\ell_1(X)}$$
3. *For every* $r > 0$, $S \subseteq X$, *for all* $f \in \mathbb{R}^S_0$,
$$r\|f\|_{\ell_1(S)} \leq \|f\|_{W_1(S,d_{X_r(S)})} \leq \left(r + \frac{diam(X, d_X)}{2}\right) \|f\|_{\ell_1(S)}$$

Proof.

1. This follows directly from the unit ball interpretation of the Wasserstein-1 norm, since $\frac{e_x-e_y}{d_X(x,y)}$ is by the first definition on the unit ball.
2. Let $m = \min_{x,y} d_X(x, y) > 0$. For distinct $x, y \in X$, we have
$$\max_{x,y\in X; x\neq y} \left\| \frac{e_x - e_y}{d_X(x, y)} \right\|_{\ell_1(X)} \leq \max_{x,y\in X; x\neq y} \frac{\|e_x - e_y\|_{\ell_1(X)}}{m} = \frac{2}{m}$$
since $0 < m \leq d_X(x, y)$ and $1 + 1 = 2$. Therefore $\frac{e_x-e_y}{d_X(x,y)} \in \frac{2}{m}B_{\ell_1(X)}$. These elements span K_{X,d_X}, so we have $K_{X,d_X} \subseteq \frac{2}{m}B_{\ell_1(X)}$ and we get the first inequality. The second inequality follows from
$$\|f\|_{W_1(X,d_X)} = \inf_{\pi\in\Pi(f^+,f^-)} \sum_{x,y\in X} d_X(x, y)\pi(x, y) \leq \mathrm{diam}(X, d_X) \sum_{x\in X} f^+(x) = \mathrm{diam}(X, d_X)\|f\|_{\ell_1(X)}/2$$
3. This inequality is a special case of the previous inequality. We have that for $X_r(S)$, $m \geq 2r$ (so $\frac{2}{m} \leq \frac{1}{r}$) and $\frac{1}{2}\mathrm{diam}(X, d_{X_r(S)}) \leq \frac{1}{2}\left(2r + \mathrm{diam}(X, d_X)\right) = \left(r + \frac{\mathrm{diam}(X,d_X)}{2}\right)$. Plugging in these estimates give the inequality.

□

4 Graph theoretic lemmas

4.1 Expanders

We will need several properties of edge-expanders in our proof of the main theorem. For this section, we fix $n, d \geq 3$ and let G be a connected n-vertex d-regular graph. We can imagine $d = 3$ in this section, all that matters is that d is fixed.

First we record two basic average bounds on distance in the shortest-path metric, denoted d_G.

Lemma 5.4.1. *Average shortest-path metric and r-magnified average shortest-path bounds.*

1. *d_G lower bound: For nonempty $S \subseteq V_G$,*

$$\frac{1}{|S|^2} \sum_{x,y \in S} d_G(x, y) \geq \frac{\log |S|}{4 \log d}$$

2. *$d_{G_r(S)}$ equality: For some $S \subseteq V_G$ and $r > 0$,*

$$\frac{1}{|E_G|} \sum_{(x,y) \in E_G} d_{G_r(x,y)} = 1 + \frac{2r|S|}{n}$$

Proof. 1. The smallest nonzero distance in G is at least 1. Thus, the average is bounded below by $\frac{1}{|S|^2}|S|(|S|-1) = 1 - \frac{1}{|S|}$ since G is connected (shortest case is complete graph on n vertices). Then, $1 - 1/a \geq (\log a)/4 \log 3$ for $a \in [15]$ ($d = 3$ maximizes), so we proceed assuming $|S| \geq 16$. Let's bound the distance in the average. Since G is d-regular, for every $x \in V_G$ the number of vertices y such that $d_G(x, y) \leq k - 1$ is at most $\sum_{i=0}^{k-1} d^i$. The rest of the vertices are farther away. Since $1 + \cdots + d^{k-2} < d^{k-1}$ we have $\#\{y : d_G(x, y) \leq k - 1\} \leq 2d^{k-1}$. Choosing $k = 1 + \lfloor \log_d(|S|/4) \rfloor$ gives that $2d^{k-1} \leq \frac{|S|}{2}$. Therefore

$$\frac{1}{|S|^2} \sum_{x,y \in S} d_G(x, y) \geq \frac{1}{|S|^2} * |S| * |S|/2 * (k - 1) = \frac{k - 1}{2} = \frac{\log(|S|/4)}{2 \log d} \geq \frac{\log |S|}{4 \log d}$$

since $|S| \geq 16$.

2. Let E_1 be edges completely contained in S and E_2 be edges partially contained in S. Because G is d-regular, $2|E_1| + |E_2| = d|S|$ (2 vertices in S for E_1, only 1 vertex for E_2, then divide by d for overcounting since each vertex in S hits d other vertices, and we count exactly the edges which have at least one vertex in S). Note that $|E_G| = dn/2$ by double-counting vertices. Then for each edge in E_1 we add $2r$, for each edge in E_2 we

add r, and otherwise we add 0 to the base distance of an edge, which is 1. Therefore,

$$\frac{1}{|E_G|}\sum_{(x,y)\in E_G} d_{G_r(x,y)} = \frac{((0+1)|E_G\setminus(E_1\cup E_2)|+(r+1)|E_2|+(2r+1)|E_1|)}{|E_G|}$$
$$= 1+\frac{r(2|E_1|+|E_2|)}{|E_G|} = 1+\frac{rd|S|}{dn/2} = 1+\frac{2r|S|}{n}$$

□

Now we introduce the definition of edge expansion.

Definition 5.4.2: Edge expansion $\phi(G)$.
G is a connected n-vertex d-regular graph. Consider $S,T\subseteq V_G$ disjoint subsets. Let $E_G(S,T)\subseteq E_G$ denote the set of edges which bridge S and T. Then the **edge-expansion** $\phi(G)$ is defined by

$$\phi(G)=\sup\left\{\phi : |E_G(S,V_G\setminus S)|\geq \phi\frac{|S|(n-|S|)}{n^2}|E_G|, \forall S\subseteq V_G, \phi\in[0,\infty)\right\}$$

We give an equivalent formulation of edge expansion via the cut-cone decomposition:

Lemma 5.4.3. *Edge-Expansion: Cut-cone Decomposition of Subsets of ℓ_1.*
$\phi(G)$ is the largest ϕ such that for all $h: V_G\to\ell_1$,

$$\frac{\phi}{n^2}\sum_{x,y\in V_G}\|h(x)-h(y)\|_1\leq\frac{1}{|E_G|}\sum_{(x,y)\in E_G}\|h(x)-h(y)\|_1$$

Proof. We will assume this for this talk. □

Now we combine Lemma 7 and the cut-cone decomposition to get Lemma 8:

Lemma 5.4.4. *Lemma* 8.
Fix $n\in\mathbb{N}$ and $\phi\in(0,1]$. Suppose G is an n-vertex graph with edge expansion $\phi(G)\geq\phi$. For all nonempty $S\subseteq V_G$ and $r>0$, every $F: V_G\to\mathbb{R}_0^S$ satisfies

$$\frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}\leq\frac{2r+{}_{diam}(S,d_G)}{(2r+1)\phi}\cdot\frac{1}{|E_G|}\sum_{(x,y)\in E_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}$$

Basically, whereas before we were bounding average norms (normal and r-magnified) in the domain space V_G, we are now bounding average norms in the image space using the Wasserstein-1 norm induced by the r-magnification of the shortest path metric on the graph.

Proof. First, plug in $h=F$ into the cut-cone decomposition, where the norm is now defined by $\ell_1(S)$. Then, we know that $\mathrm{diam}(S,d_{G_r}(S))=2r+{}_{\mathrm{diam}}(S,d_G)$ and the smallest positive distance in $(S,d_{G_r(S)})$ is $2r+1$. Applying Lemma 7, every $x,y\in V_G$ satisfy

$$\frac{2r+1}{2}\|F(x)-F(y)\|_{\ell_1(S)}\leq\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}\leq\frac{2r+{}_{\mathrm{diam}}(S,d_G)}{2}\|F(x)-F(y)\|_{\ell_1(S)}$$

Plugging these estimates directly into the cut-cone decomposition

$$\frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{\ell_1(S)} \leq \frac{1}{\phi|E_G|}\sum_{(x,y)\in E_G}\|F(x)-F(y)\|_{\ell_1(S)}$$

gives the desired result. □

4.2 An Application of Menger's Theorem

Here, we want to bound below the number of edge-disjoint paths.

Lemma 5.4.5. *Lemma* 9.
Let G be an n-vertex graph and $A, B \subseteq V_G$ be disjoint. Fix $\phi \in (0,\infty)$ and suppose $\phi(G) \geq \phi$. Then

$$\#\left\{\textit{edge-disjoint paths joining} A \textit{ and} B\right\} \geq \frac{\phi|E_G|}{2n}\cdot\min\{|A|,|B|\}$$

Proof. Let m be the maximal number of edge-disjoint paths joining A and B. By Menger's theorem from classical graph theory, there exists a subset of edges $E^* \subseteq E_G$ with $|E^*| = m$ such that every path in G joining $a \in A$ to $b \in B$ contains an edge from E^*.

Now consider the graph $G^* = (V_G, E_G \setminus E^*)$. In this graph, there are no paths between A and B. Now, if we let $C \subseteq V_G$ be the union of all connected components of G^* containing an element of A, then $A \subseteq C$ and $B \cap C = \emptyset$. Since we covered the maximal possible vertices reachable from A with C, all edges between C and $V_G \setminus C$ are in E^*. Therefore, $|E_G(C, V_G \setminus C)| \leq |E^*| = m$. By the definition of expansion,

$$m \geq |E_G(C, V_G\setminus C)| \geq \phi\frac{\max\{|C|, n-|C|\}\cdot\min\{|C|, n-|C|\}}{n^2}|E_G| \geq \frac{\phi\min\{|A|,|B|\}\cdot|E_G|}{2n}$$

since $\max\{|C|, n-|C|\} \geq n/2$ and since $A \subseteq C, B \subseteq V_G \setminus C$, we have $\min\{|C|, n-|C|\} \geq \min\{|A|,|B|\}$.

□

5 Main Proof

We fix $d, n \in \mathbb{N}, \phi \in (0,1)$ and let G be a d-regular graph on n vertices with $\phi(G) \geq \phi$. We also fix a nonempty subset $S \subset V_G$ and $r > 0$ and define a mapping

$$f : (S, d_{G_r(S)}) \mapsto \left(\mathbb{R}^S_0, \|\cdot\|_{W_1(S, d_{G_r(S)})}\right) \quad \text{s.t.} \quad f(x) = e_x - \frac{1}{|S|}\sum_{z\in S} e_z \ \forall \ x \in S$$

Then suppose we have that some $F : V_G \mapsto \mathbb{R}^S_0$ extends f and for some $L \in (0,\infty)$ we have

$$\|F(x)-F(y)\|_{W_1(S, d_{G_r(S)})} \leq L d_{G_r(S)}(x,y)$$

For $x \in V_G, s \in (0,\infty)$ define

$$B_S(x) = \{y \in V_G : \|F(x) - F(y)\|_{W_1(S,d_{G_r(S)})} \leq s\}$$

i.e.e B_s is the inverse image of F of the ball (in the Wasserstein 1-norm) of radius s centered at $F(x)$. By the lemma consequence of Menger's Theorem, since $\phi \geq \phi(G)$ we have

$$m \geq \frac{\phi d}{4} \min\{|S \backslash B_s(x)|, |B_s(x)|\} \qquad (1)$$

for m edge-disjoint paths between $S\backslash B_s(x)$ and $B_s(x)$, i.e. we can find indices $k_1, \ldots, k_m \in \mathbb{N}$ and vertex sets $\{z_{j,1}, \ldots, z_{j,k_j}\}_{j=1}^m \in V_G$ s.t. $\{z_{j,1}\}_{j=1}^m \subset S\backslash B_s(x)$, $\{z_{j,k_j}\}_{j=1}^m \subset B_s(x)$ (i.e. the beginnings and ends of paths are in different disjoint subsets) and such that $\{\{z_{j,1}, z_{j,i+1} : j \in \{1, \ldots, m\} \wedge i \in \{1, \ldots, k_j - 1\}\}$ are distinct edges in E_G (i.e. edge-disjointedness). Now take an index subset $J \subset \{1, \ldots, m\}$ s.t. $\{z_{j,1}\}_{j\in J}$ are distinct and $\{z_{j,1}\}_{j \in J} = \{z_{i,1}\}_{i=1}^m$. For $j \in J$ denote by d_j the number of $i \in \{1, \ldots, m\}$ for which $z_{j,1} = z_{i,1}$. Then since G is d-regular and $\{\{z_{i,1}, z_{i,2}\}\}_{i=1}^m$ are distinct, $\max\limits_{j\in J} d_j \leq d$. Since $\sum\limits_{j\in J} d_j = m$, from (1) we have that

$$|J| \geq \frac{m}{d} \geq \frac{\phi}{4} \min\{|S\backslash B_s(x)|, |B_s(x)|\} \qquad (2)$$

The quantity $|J|$ can be upperbounded as follows:

Lemma 5.5.1. *Lemma* 10*:* $|J| \leq \max\left\{d^{16(s-r)}, \frac{16nLd\log d}{\log n}\left(1 + \frac{2r|S|}{n}\right)\right\}$

Proof. Assume $|J| \leq d^{16(s-r)}$ (otherwise we are done). This is equivalent to

$$s - r < \frac{\log |J|}{16 \log d}$$

Now since $\{z_{j,1}\}_{j\in J} \subset S$ and $F(x) = f(x)\ \forall\ x \in S$ and is an isometry on $(S, d_{G_r(S)})$, by the definition of the r-magnified metric

$$\|F(z_{i,1}) - F(z_{j,1})\|_{W_1(S,d_{G_r(S)})} = d_{G_r(S)}(z_{i,1}, z_{j,1}) = 2r + d_G(z_{i,1}, z_{j,1})$$

This gives us

$$\begin{aligned}\sum_{j\in J} \|F(z_{j,1}) - F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})} &= \frac{1}{2|J|}\sum_{i,j\in J}\left(\|F(z_{i,1}) - F(z_{i,k_i})\|_{W_1(S,d_{G_r(S)})} + \|F(z_{j,1}) - F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})}\right)\\ &\geq \frac{1}{2|J|}\sum_{i,j\in J}\left(\|F(z_{i,1}) - F(z_{j,1})\|_{W_1(S,d_{G_r(S)})} + \|F(z_{z_i,k_i}) - F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})}\right)\end{aligned}$$

Since $\{z_{j,k_j}\}_{j\in J} \subset B_s(x)$, by the definition of $B_s(x)$ we have

$$\|F(z_{i,k_i}) - F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})} \leq \|F(z_{i,k_i}) - F(x)\|_{W_1(S,d_{G_r(S)})} + \|F(x) - F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})} \leq 2s\ \ \forall\ \ i,j \in J$$

so the previous inequality can be further continued as

$$\sum_{j\in J}\|F(z_{j,1})-F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})} \geq \frac{1}{2|J|}\sum_{i,j\in J} d_G(z_{i,1},z_{j,1})-(s-r)|J| \geq \frac{|J|\log|J|}{8\log d}-(s-r)|J| > \frac{|J|\log|J|}{16\log d}$$

where the second inequality above is a property of the expander graph. The same quantity can be bounded from above using the Lipschitz condition

$$\sum_{j\in J}\|F(z_{j,1})-F(z_{j,k_j})\|_{W_1(S,d_{G_r(S)})} \leq L\sum_{j\in J} d_{G_r(S)}(z_{j,1},z_{j,k_j}) \leq L\sum_{j\in J}\sum_{i=1}^{k_j-1} d_{G_r(S)}(z_{j,1},z_{j,i+1})$$

By the edge-disjointedness of the paths (specifically since $\{\{z_{j,1},z_{j,i+1}\} : j \in J \wedge i \in \{1,\ldots,k_j-1\}\})$ are distinct edges in E_G, we have that

$$\sum_{j\in J}\sum_{i=1}^{k_j-1} d_{G_r(S)}(z_{j,1},z_{j,i+1}) \leq \sum_{\{u,v\}\in E_G} d_{G_r(S)}(u,v) = \frac{nd}{2}\left(1+\frac{2r|S|}{n}\right)$$

Everything together gives

$$\frac{Lnd}{2}\left(1+\frac{2r|S|}{n}\right) \geq \frac{|J|\log|J|}{16\log d}$$

By the simple fact that $a\log a \leq b \implies a \leq \frac{2b}{\log b}$ for $a \in [1,\infty), b \in (1,\infty)$, using $a = |J|$ and $b = 8Lnd\log d\left(1+\frac{2r|S|}{n}\right) \geq n$ we have that

$$|J| \leq \frac{16nLd\log d}{\log n}\left(1+\frac{2r|S|}{n}\right)$$

completing the proof. □

The lemma has two corollaries, both depending on the following condition:

$$d^{16(s-r)} \leq \frac{\phi|S|}{8} \qquad L \leq \frac{\phi|S|\log n}{128\left(1+\frac{2r|S|}{n}\right)nd\log d} \qquad (3)$$

Corollary 5.5.2. *Corollary* 11*:* $\max_{x\in V_G}|B_s(x)| < \frac{|S|}{2}$

Proof. Pick an $x \in V_G$. If $B_s(x)\cap S$ is nonempty, then again using the standard estimate on expander graphs we have that $\exists\ y,z \in B_s(x)\cap S$ s.t.

$$d_G(y,z) \geq \frac{\log|B_s(x)\cap S|}{4\log d}$$

Since $y,z \in S$ and $F(x) = f(x)\ \forall\ x \in S$ and furthermore F is an isometry on $(S, d_{G_r(S)})$, we have similarly to before that

$$d_G(y,z)+2r = d_{G_r(S)}(y,z) = \|F(y)-F(z)\|_{W_1(S,d_{G_r(S)})} \leq \|F(y)-F(x)\|_{W_1(S,d_{G_r(S)})}+\|F(x)-F(z)\|_{W_1(S,d_{G_r(S)})}$$

where we have used first the triangle inequality and second the fact that $y, z \in B_s(x)$. Then using the first inequality in the proof we have

$$|B_s(x) \cap S| \leq d^{8(s-r)} \leq \sqrt{\frac{\phi|S|}{8}} \leq \frac{2|S|}{5}$$

where the second inequality comes from the first assumption. Now the above inequality implies that $|S \backslash B_s(x)| \geq \frac{3|S|}{5}$, which combined with Lemma 10 yields

$$\min\left\{\frac{3|S|}{5}, |B_s(x)|\right\} < \max\left\{\frac{4d^{16(s-r)}}{\phi}, \frac{64nLd\log d}{\phi \log n}\left(1 + \frac{2r|S|}{n}\right)\right\}$$

However, the two original assumptions together imply that $\frac{3|S|}{5}$ is in fact greater than either value in the maximum, so

$$|B_s(x)| \leq \max\left\{\frac{4d^{16(s-r)}}{\phi}, \frac{64nLd\log d}{\phi \log n}\left(1 + \frac{2r|S|}{n}\right)\right\} \leq \frac{|S|}{2}$$

$\square$

Corollary 5.5.3. *Corollary* 12: $L \geq \frac{\phi s}{2\left(1+\frac{diam(G,d_G)}{2r}\right)\left(1+\frac{2r|S|}{n}\right)}$

Proof. By the definition of $B_s(x)$ for $x \in V_G$ and $y \in V_G \backslash B_s(x)$ we have $\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} > s$, so

$$\begin{aligned}\frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} &\geq \frac{1}{n^2}\sum_{x\in V_G}\sum_{y\in V_G\backslash B_s(x)}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}\\ &\geq \frac{s}{n^2}\sum_{x\in V_G}(n-|B_s(x)|) \geq s\left(1-\frac{\max_{x\in V_G}|B_s(x)|}{n}\right) \geq \frac{s}{2}\end{aligned}$$

where the last inequality comes from Corollary 11. Then since Lemma 8 gives

$$\begin{aligned}\frac{1}{n^2}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})} &\leq \frac{2r+\mathrm{diam}(S,d_G)}{(2r+1)\phi|E_G|}\sum_{x,y\in V_G}\|F(x)-F(y)\|_{W_1(S,d_{G_r(S)})}\\ &\leq \frac{L\left(1+\frac{\mathrm{diam}(G,d_G)}{2r}\right)}{\phi|E_G|}\sum_{\{x,y\}\in E_G} d_{G_r(S)}(x,y)\\ &= \frac{L\left(1+\frac{\mathrm{diam}(G,d_G)}{2r}\right)\left(1+\frac{2r|S|}{n}\right)}{\phi}\end{aligned}$$

where the ineqaulity on the second line is true since F is an isometry on (S, d_G) and from the Lipschitz constant of f and the equality is average length of the r-magnification of the expander graph G. Combining these inequalities with the previous ones yields the corollary. $\square$

Theorem 5.5.4. *Theorem* 13*: if* $0 < r \le diam(G, d_G)$ *then*

$$L \geq_C \frac{\phi}{1+\frac{r|S|}{n}} \min\left\{\frac{|S|\log n}{nd\log d}, \frac{16r^2\log d + r\log\left(\frac{\phi|S|}{8}\right)}{diam(G,d_G)\log d}\right\}$$

Proof. Assume $16r\log d + \log\left(\frac{\phi|S|}{8}\right) > 0$ (otherwise we are done) and choose $s = r + \frac{\log\left(\frac{\phi|S|}{8}\right)}{16\log d}$ s.t. $s > 0$ and $d^{16(s-r)} = \frac{\phi|S|}{8}$. Then the first inequality in (3) is satisfied, so either the second fails and L thus has that expression as a lower bound, or both are satisfied and we have the lower bound in Corollary 12. $\square$

Theorem 5.5.5. *Theorem* 1*: for every* $n \in \mathbb{N}$ *we have* $\boldsymbol{ae}(n) \geq_C \sqrt{\log n}$.

Proof. Substituting $\phi \asymp 1$ and $\mathrm{diam}(G, d_G) \asymp \frac{\log n}{\log d}$ (the $\asymp$ indicates asymptotically for large n), and using the assumption $0 < r \le \mathrm{diam}(G, d_G)$, the lower bound given in Theorem 13 becomes

$$L \geq_C \frac{1}{1+\frac{r|S|}{n}} \min\left\{\frac{|S|\log n}{nd\log d}, \frac{r(r\log d + \log|S|)}{\log n}\right\}$$

Taking $S \subset V_G$ s.t. $|S| = \left\lfloor \frac{n\sqrt{d\log d}}{\sqrt{\log n}} \right\rfloor$ (we must have n

ged^d to ensure $|S| \le n$) and $r \asymp \frac{\log d}{\sqrt{d\log d}}$, which gives a lower bound of $L \geq_C \frac{\sqrt{\log n}}{\sqrt{d\log d}}$. Therefore

$$e_{\left\lfloor \frac{n\sqrt{d\log d}}{\sqrt{\log n}} \right\rfloor}(G_r(S), W_1(S, d_{G_r(S)})) \geq_C \frac{\sqrt{\log n}}{\sqrt{d\log d}}$$

which completes the proof for fixed d. $\square$

List of Notations

Index