Metric embeddings and geometric inequalities

taught by Assaf Naor

February 8, 2016

Contents

Introduction

Notes from Assaf Naor's class "Geometric inequalities with applications to metric embeddings" at Princeton in Spring 2016.

More generally we'll cover metric geometry.

I'll teach things that just appeared or are still in the works. Most of them aren't in books yet; some of them haven't even be written yet. I'll give a taste of what's going on in the field. I'll choose topics with many open questions associated with them and with self-contained proofs.

This is a huge area. I'll use proofs of major theorems as excuses to see the lemmas that go into the proofs. The theorems are famous; in addition to being famous, they use lemmas that are immensely powerful. Instead of presenting the geometric inequalities and then applications afterwards, we'll see the lemmas on the way to proving these big theorems.

Express your interest—I'm flexible.

Chapter 1

The Ribe Program

2/1/16

1 The Ribe Program

Our main motivation is the **Ribe Program**. The program is inspired from a theorem from 1975 called Ribe's Rigidity Theorem 1.1.8. The theorem is about Banach spaces, specifically a relationship between their linear structure and metric structure.

Definition 1.1.1: Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ by Banach spaces. We say that X is (crudely) **finitely representable** in Y if there exists $K > 0$ 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 1.1.2: 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 1.1.3: 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 $F = \text{span}\{x_1, \ldots, x_n\}$. 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 1.1.4 (Kahane's inequality)**.** *For any normed space Y and $q \ge 1$, for all n, $y_1, \ldots, y_n \in Y$,*

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

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 1.1.2 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?

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

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 1.1.5: 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 ($\forall x_1, \ldots, x_n, P(x_1, \ldots, x_n)$ 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 1.1.6: 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 1.1.7 (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 1.1.8 (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 1.1.9. *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 1.1.10: 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 the theorem, people wondered: can we reformulate the local theory of Banach spaces without mentioning anything about the linear structure? In metric spaces, we are only allowed to discuss distances between points, not linear properties (ex. summing up). Suppose we can reformulate local theory in this way—find a dictionary that reformulates each linear property and theorem about linear properties as properties and theorems involving distances. Then we can state the analogous theorems for metric spaces. In particular, we can discuss when metric spaces have type and cotype. Maybe the theorems remain true—often they do, for different reasons. Now we can apply the theorem to graphs, groups, etc. 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 1.1.11. 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 1.1.8 and Proposition 1.1.3. Can we prove something like this theorem without using Ribe's Theorem 1.1.8? 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, $x_1, \ldots, x_n \in X$. 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

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

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

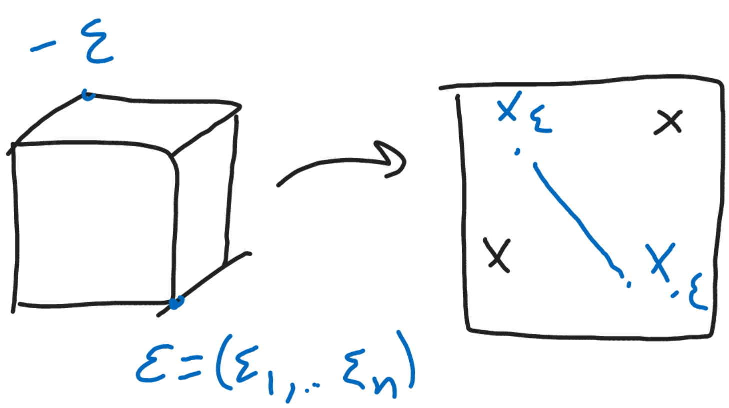
eq:type-gen
$$\mathbb{E}\left[\|f(\varepsilon) - f(-\varepsilon)\|^p\right] \lesssim_X \sum_{i=1}^n \mathbb{E}\left[\|f(\varepsilon) - f(\varepsilon_1, \ldots, \varepsilon_{i-1}, -\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_n)\|^p\right]. \tag{1.2}$$

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

Definition 1.1.12: df:enflo A metric space (X, d_X) has **Enflo type** p if there exists $T > 0$ such that for every n and every $f : \{\pm 1\}^n \to X$,

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

This is bold. In (1.2), the points had to be vertices of a cube, but in Definition 1.1.12, they are arbitrary. The moment you choose the labelings, you impose a cube structure between the points.



The inequality says:

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

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 1.1.13 (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 1.1.14 (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.

The conjecture is true for many spaces. For example, it's true for L^4. 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.

The moment that I throw away the linear structure, Corollary 1.1.11 becomes easy.

Proposition 1.1.15: pr:uh-enflo If X, Y are uniformly homeomorphic Banach spaces and Y has Enflo type p, then so does X.

Lemma 1.1.16 (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 1.1.15 given Lemma 1.1.16. By definition of uniformly homeomorphic, there exists a homeomorphism $\psi : X \to Y$ such that ψ, ψ^{-1} are uniformly continuous. Lemma 1.1.16 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. □

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 1.1.16.

Proof of Lemma 1.1.16. 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

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

hen

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

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 1.2.3. There is an amazing open problem in this context.

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

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

(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 1.2.3 says that you can choose $\delta = \delta_n$ to be independent of the geometry of X and Y.

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.

Definition 1.2.1: Suppose $(X, d_X), (Y, d_Y)$ are metric spaces $D \geq 1$. We say that X embeds into Y with **distortion** D if there exists $f : X \to Y$ and $s > 0$ such that for all $x, y \in X$,

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

The infimum over those $D \geq 1$ such that X embeds into Y with distortion is denoted $C_Y(X)$.

Definition 1.2.2: 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 $B_X := \{x \in X : \|x\| \leq 1\}$.

The way to read this is that if you got $C_Y(\mathcal{N}_\delta)$ to be small, then the distortion of the entire object is not much larger.

Theorem 1.2.3 (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 $C_Y(X) \leq 2C_Y(\mathcal{N}_\delta)$.

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 1.2.4: 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

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

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

Theorem 1.2.5 (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 1.1.8.

Proof of Ribe's Theorem 1.1.8 from Bourgain's Theorem 1.2.3. Let X, Y be Banach spaces that are uniformly homeomorphic. By Corson-Klee 1.1.16, 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 1.2.3, 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. $\square$

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 1.2.3.

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

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'.*

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. Von Neumann wrote a paper called "Matric spaces," and the only thing he proves is this inequality, and he didn't know what it was useful for. Schatten later wrote books on applications. There is a lot to do with this!

Hint: it's short but it's not typical linear algebra!

Spielman and Srivastava have a beautiful theorem.

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

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

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 $|\sigma|\geq(1-\varepsilon)srank(A)$ such that*

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

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

This is stronger than Bourgain-Tzafriri. To obtain Bourgain-Tzafriri the columns were unit vectors. Then $\|A\|_{S_2}=\sqrt{m}$ (why) and $\mathrm{srank}(A)=\frac{m}{\|A\|_{S_\infty}^2}$. 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.

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 $k<\mathrm{rank}(A)$ then there exist $\sigma\subseteq\{1,\ldots,m\}$ with $|\sigma|=k$ such that*

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

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 as big as you want up to the rank, and we have sharp control in the entire range.

Next time we'll show how this implies the other theorems, and then prove the theorem.
2-8-16

A subsequent theorem gave the same theorem but instead of the stable rank, used something bigger.

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 $\sigma\subseteq\{1,\ldots,m\}$, $|\sigma|=k$ such that

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

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

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 $p \ge q > 2$, then

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

(Hint: Use Hölder's inequality.)

Proof of Theorem 2.1.7. Note that

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

□

To optimize, fixing 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.

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

Lemma 2.1.11 (Ky Fan maximum principle)**:** Suppose that $P : \mathbb{R}^m \to \mathbb{R}^m$ is a rank k orthogonal projection. Then

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

Proof. This proof isn't complete. I will fix it next time.

We will prove that if $B : \mathbb{R}^m \to \mathbb{R}^m$ is positive semidefinite, then

$$\operatorname{Tr}(BP) \le \sum_{i=1}^k s_i(B).$$

To get the lemma, set $B = A^*A$.

Apply arbitrarily small perturbation s_i so that

$$s_1(B) > s_2(B) > \cdots > s_m(B).$$

Let $v_1, \ldots, v_m$ be an orthonormal basis such that $Bv_i = s_i(B)v_i$. Let $u_1, \ldots, u_k$ be an orthonormal basis of $P\mathbb{R}^m$ ordered so that

$$\langle Bu_1, u\rangle \ge \langle Bu_2, u_2\rangle \ge \cdots \ge \langle Bu_k, u_k\rangle .$$

We calculate

$$\begin{aligned}\operatorname{Tr}(BP) &= \sum_{i=1}^k \langle BPu_i, u_i\rangle \\ &= \sum_{i=1}^k \langle Bu_i, u_i\rangle .\end{aligned}$$

We will prove by induction on i that

$$\langle Bu_i, u_i\rangle \le s_i(B).$$

For $i = 1$,

$$s_1(B) = \|B\|_{S_\infty},$$

and (when $\|u_1\| = 1$)

$$\langle Bu_1, u_1\rangle \le \|B\|_{S_\infty} = s_1(B).$$

Suppose we proved the claim for $j-1$. If ?$\langle Bu_i, u_i\rangle \le s_j(B)$ $\langle Bu_{j-1}, u_{j-1}\rangle \le s_j(B)$ then we're done because $\langle Bu_j, u_j\rangle \le \langle Bu_{j-1}, u_{j-1}\rangle \le s_j(B)$. So we may assume that $\langle Bu_{j-1}, u_{j-1}\rangle > s_j(B)$.

Write the u's in the basis of v's:

$$u_i = \sum_{l=1}^m c_{il} v_l.$$

The fact that the u_i's are orthonormal means that the c_i's are probability vectors,

$$\sum_{l=1}^{m} c_{il}^2 = 1.$$

We have

$$\langle Bu_i, u_i \rangle = \sum_{l=1}^{m} m c_{il}^2 s_l(B).$$

If $c_{il}^2 > 0$ for any $l \geq j$. $\square$

1.1 Finding big subsets

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

Theorem 2.1.12 (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 $x_1, \ldots, x_k \in \mathbb{R}^m$,*

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

for some $c \in \{1, \ldots, m\}$ 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?

$$\|Tx_1\|_2 \leq \sqrt{\frac{\pi}{2}} \|T\|_{\ell_\infty^m \to \ell_2^n} \|X_1\|_\infty$$

We know the inequality for $k = 1$ with constant 1, and we get it true for arbitrary many vectors, losing an universal constant $(\left(\frac{\pi}{2}\right))$. 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 2.1.13. clm:lgi

$$\text{eq:lgi1} \sum_{j=1}^{m} \left(\sum_{r=1}^{k} (T^*Tx_r)_j^2 \right)^{\frac{1}{2}} \leq \sqrt{\frac{\pi}{2}} \|T\|_{\ell_\infty^m \to \ell_2^n} \left(\sum_{r=1}^{k} \|Tx_r\|^2 \right)^{\frac{1}{2}}. \tag{2.1}$$

Proof of Theorem 2.1.12. Assuming the claim, we prove the theorem.

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

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 2.1.13. Let $g_1, \dots, g_k$ be iid standard Gaussian random variables. For every fixed $j \in \{1, \dots, m\}$,

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

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

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

Sum these over j:

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

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

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

Then

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

This is a pointwise inequality. Taking expectations,

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

What is the second moment? Expand:

$$\mathbb{E}\left\|\sum_{r=1}^k g_i T x_r\right\|_2^2 = \sum_{r=1}^k \|T x_r\|_2^2 = \mathbb{E}\left[\sum_{ij} g_i g_j \langle T x_i, T x_j \rangle\right].$$

$\square$

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. Some fact about restricting matrices. Another geometric argument to give a different method for selecting subsets. Combinatorial lemma for selecting subsets. Put together in crazy induction.

From this proof you can reverse engineer vectors that make the inequality sharp. You need to come up with T and the points.

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

$$Tf = (\mathbb{E}[f g_1], \ldots, \mathbb{E}[f g_m]).$$

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

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

Index