Local Theory of Banach Spaces*

Naor, Fall 2010

Scribe: Evan Chou

Texts:

- Milman, Schechtman. *Asymptotic theory of finite dimensional normed spaces*
- Albiac, Kalton. *Topics in Banach space theory*
- Pisier. *Volumes of convex bodies and Banach space geometry*
- Tomczak, Jaegerman. *Banach-Mazur distances and finite dimensional operator ideals*
- Ball. *Flavors of Geometry: An elementary introduction to modern convex geometry.*

Table of contents

*. This document has been written using the GNU $\text{T}_\text{E}\text{X}_\text{MACS}$ text editor (see `www.texmacs.org`).

Week 1 (9/7/2010)

Overview

What do we mean by **Local Theory**? We care about structures of infinite dimensional Banach spaces, for example:

- $L_p(\mu)$
- $C(K)$ for K compact, Hausdorff
- l_p
- $l_p^n := (\mathbb{R}^n, \|\cdot\|_p)$
- c_0, space of null sequences (limit as $n \to \infty$ is 0)

It turns out that many global properties of Banach spaces can be learned from properties of finite dimensional substructures, and this is what we mean by "local". Of course, this means that we have to start caring about finite dimensional spaces.

In finite dimensional spaces, all norms on $\mathbb{R}^n$ are equivalent. If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, we can associate the norm to the unit ball

$$B_{\|\cdot\|} = \{x \in \mathbb{R}^n : \|x\| \le 1\}$$

Denoting $K = B_{\|\cdot\|}$, we know that K is convex and symmetric ($x \in K \Longrightarrow -x \in K$). Conversely, if we have some subset $K \subset \mathbb{R}^n$ which is convex, symmetric, and has nonempty interior, then we can define a norm

$$\|x\|_K := \inf\left\{\lambda > 0 : \frac{x}{\lambda} \in K\right\}$$

for which $K = B_{\|\cdot\|_K}$.

Thus, studying norms in $\mathbb{R}^n$ is equivalent to studying convex bodies. For instance, for l_p^3, the unit ball for $p = 2$ is the sphere, the unit ball for $p = \infty$ is the unit cube, and the unit ball for $p = 1$ is a tetrahedron. The ball in l_1^3 has sharp edges (pointy), but in higher dimensions the ball in l_1^n is harder to visualize, and in fact this pointy intuition is not correct:

Theorem 1. **(Figiel-Lindenstrauss-Milman)** *For all $\varepsilon > 0$, there exists $c(\varepsilon) > 0$ such that for all n, there exists a linear subspace $F \subset \mathbb{R}^n$ and $r > 0$ such that*

1. $r(B_{l_2^n} \cap F) \subset (F \cap B_{l_1^n}) \subset (1 + \varepsilon) r(B_{l_2^n} \cap F)$
2. $\dim F \geq c(\varepsilon) n$

In words, there exists a lower dimensional slice of $B_{l_1^n}$ (which is supposedly pointy) wedged between slices of two Euclidean balls (round) with arbitrarily close radii.

For l_∞^n the theorem is the same with $c(\varepsilon)n$ replaced by $c(\varepsilon) \log n$, and it can be shown that l_∞^n is the worst in some sense (that if the above property holds, then $\dim F \lesssim C_2(\varepsilon) \log n$ also).

Next, consider a normed space $(X, \|\cdot\|)$, with n unit vectors $x_1, ..., x_n \in X$, $\|x_i\| = 1$. We have the triangle inequality

$$\|\varepsilon_1 x_1 + ... + \varepsilon_n x_n\| \leq n$$

for all choice of signs $\varepsilon_i \in \{\pm 1\}$. Also,

$$\mathbb{E}_\varepsilon \|\varepsilon_1 x_1 + ... + \varepsilon_n x_n\| \leq n$$

for some probability distribution over $\varepsilon_i \in \pm 1$. This inequality is sharp in the case when $X = l_1^n$, $x_i = e_i$. We make the remark that if ε_i is uniform ± 1, and all $x_i = x$ are identical, then this corresponds to a random walk, with expectation bounded above by $\sqrt{n}$, and without the presence of the sign terms ε_i, the inequality can be achieved by setting all $x_i = x$, but this is a case to rule out.

We will prove (eventually) that l_1^n is the only obstruction for improving this inequality:

Theorem 2. **(Pisier)** *For all $(X, \|\cdot\|)$, one of the following holds:*

1. *There exists $\alpha < 1$ and constant K, such that for all n and $x_1, ..., x_n \in X$ unit vectors,*

$$\mathbb{E}\left[\left\|\sum_i \varepsilon_i x_i\right\|\right] \leq K n^\alpha$$

2. *Or, for all n and $\varepsilon > 0$, there exists a linear operator $T: l_1^n \to X$ with $\|x\| \leq \|Tx\| \leq (1 + \varepsilon)\|x\|$ (in which case the bound cannot be improved from $(1 \pm \varepsilon)n$)*

This says that if you cannot improve triangle inequality bound, then there is a l_1^n lurking inside the space. Now we turn to basic results that we will be using throughout the course.

Basic Results and Tools

Existence of Haar Measure

Theorem 3. *Let (M, d) be a compact metric space, G a group acting on M by isometries, i.e.*

$$d(gx, gy) = d(x, y) \text{ for all } x, y \in M \text{ and } g \in G$$

Then there exists a regular measure on the Borel sets which is G-invariant, i.e.

$$\mu(gA) = \mu(A) \text{ for all Borel sets } A, g \in G$$

Moreover, if G acts transitively on M ($Gx = M$ for all $x \in M$), then μ is unique up to multiplication by scalars.

Many proofs, and here is one that is particularly short and tricky.

Proof. For all $\varepsilon > 0$, let N_ε be a minimal ε-net in M, i.e. $\bigcup_{x \in N_\varepsilon} B(x, \varepsilon) = X$ and $|N_\varepsilon|$ is minimal with respect to this property. (Notation: $B(x, \varepsilon)$ will always refer to the closed ball $\{y: d(x, y) \leq \varepsilon\}$).

Look at $C(M)$, the space of continuous functions on M, and define the linear functional

$$\mu_\varepsilon(f) = \frac{1}{n_\varepsilon} \sum_{x \in N_\varepsilon} f(x)$$

In addition to being linear, $\mu_\varepsilon(1) = 1$, μ_ε is positive, i.e. if $f \geq 0$ then $\mu(f) \geq 0$, and also, the operator norm of μ_ε is $\|\mu_\varepsilon\| \leq 1$. We have a sequence of measures that are uniformly bounded in norm, and thus by Banach-Alaoglu weak* compactness there exists a sequence $\varepsilon_i \to 0$ and a linear functional μ on $C(M)$ such that $\mu_{\varepsilon_i} \to \mu$ in the weak* sense, i.e. for all $f \in C(M)$,

$$\mu_{\varepsilon_i}(f) \to \mu(f)$$

The limiting μ inherits the same properties, that $\mu(1) = 1$ and positivity $f \geq 0 \Longrightarrow \mu(f) \geq 0$. Now by the Riesz Representation Theorem for $C(M)^*$, μ is actually a measure on M, so we can write

$$\mu(f) = \int_M f d\mu$$

and since $\mu(1) = 1$, μ is a probability measure.

Now we have a small claim:

Claim: If $\{N'_\varepsilon\}_{\varepsilon > 0}$ is any minimal ε-net, and we define a corresponding $\mu'_\varepsilon(f) = \frac{1}{n_\varepsilon} \sum_{x \in N'_\varepsilon} f(x)$, then also

$$\mu'_{\varepsilon_i}(f) \to \mu(f) \text{ for all } f$$

as well. In other words, the limiting μ is independent of the minimal ε-net that we use.

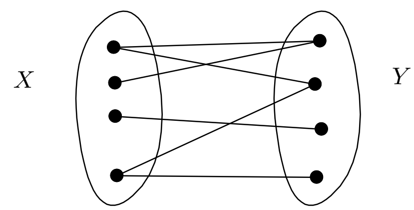
To prove this, we will show that there exists a 1-1 and onto mapping ψ: $N_\varepsilon \to N'_\varepsilon$ such that $d(x, \psi(x)) \leq 2\varepsilon$. If this is true, then

$$|\mu_\varepsilon(f) - \mu'_\varepsilon(f)| = \left| \frac{1}{n_\varepsilon} \sum_{x \in N_\varepsilon} (f(x) - f(\psi(x))) \right| \leq \sup_{d(a,b) \leq 2\varepsilon} |f(a) - f(b)| \overset{(\varepsilon \to 0)}{\longrightarrow} 0$$

noting that f is a continuous function on a compact set, and is therefore uniformly continuous, so the last quantity goes to 0 as $\varepsilon \to 0$.

To prove that we can find ψ, we can use a combinatorial result called the Hall Marriage Theorem:

Theorem 4. (Hall Marriage Theorem) *Let X, Y be sets with $|X| = |Y|$ in a bipartite graph:*



For $x \in X$ and $y \in Y$ we use $x \sim y$ to notate that x knows y (is connected to). For a subset $A \subset X$, denote $K(A)$ to be the set of people that A knows, i.e.

$$K(A) = \{y \in Y : x \sim y \text{ for some } x \in A\}$$

Then there exists a one-to-one and onto function $f: X \to Y$ such that $x \sim f(x)$ if and only if $|K(A)| \geq |A|$ for all $A \subset X$.

The proof of this theorem is by induction, trivial for the case where $|X| = |Y| = 1$. First, note that if we have that $|A| = |K(A)|$ for some subset $A \subset X$, then we can decompose this problem into two smaller problems, one with A and $K(A)$, and one with $B := X \backslash A$ and $K(B)$. Note that for $A, K(A)$, the condition is automatically satisfied. For $B, K(B)$, if the condition is not satisfied by some set S, then we note that the condition will not be satisfied by $X \cup S$ for the full graph. The only case left is if $|A| < |K(A)|$ for all A. In this case we may remove any connected pair (x, y) and apply induction to the rest of the graph.

To apply this theorem, we will use N_ε and N'_ε, and say that $x \in N_\varepsilon$ knows $y \in N'_\varepsilon$ if $B(x, \varepsilon) \cap B(y, \varepsilon)$ is nonempty. Now we show that for any $A \subset N_\varepsilon$ that $|A| \leq |K(A)|$ by contradiction. Suppose that $|A| > |K(A)|$. Then we can create a new covering with $(N_\varepsilon \backslash A) \cup K(A)$ with fewer elements, which is a contradiction. Note that $K(A)$ necessarily must cover at least the same region that A covers, since if $z \in M$ is covered by $B(x, \varepsilon)$, then there is some $y \in N'_\varepsilon$ which also covers z, in which case $z \in B(x, \varepsilon) \cap B(y, \varepsilon)$. Thus the conditions of the Hall Marriage Theorem are satisfied, and we can find a mapping $\psi: N_\varepsilon \to N'_\varepsilon$ which is one to one and onto, and moreover, $d(x, \psi(x)) \leq 2\varepsilon$ from the knowing condition.

Note that we have not used any of the group conditions so far. Now consider $g \in G$. Since G acts on M by isometries, we have that gN_ε is also a minimal ε-net. Thus, we have that

$$\mu(f) = \lim_{i \to \infty} \mu_{\varepsilon_i}(f) = \lim_{i \to \infty} \mu_{\varepsilon_i}(f \circ g) = \mu(f \circ g)$$

Thus

$$\int f(x)\, d\mu(x) = \int f(gx)\, d\mu(x) \text{ for all } g \in G, f \in C(M)$$

and so we have found a measure μ which is invariant under the action of G.

Uniqueness. To show uniqueness, first note that G inherits a metric from M:

$$\rho(g, h) := \sup_{x \in M} d(gx, hx)$$

for $g, h \in G$. This satisfies the triangle inequality:

$$\begin{aligned}
\rho(g_1, g_2) &= \sup_{x \in M} d(g_1 x, g_2 x) \\
&\leq \sup_{x \in M} d(g_1 x, hx) + d(hx, g_2 x) \\
&\leq \sup_{x \in M} d(g_1 x, hx) + \sup_{y \in M} d(hy, g_2 y) \\
&= \rho(g_1, h) + \rho(g_2, h)
\end{aligned}$$

but it may not be the case that $\rho(g,h)=0 \Longrightarrow g=h$. If not, this means that $gx=hx$ for all x, or $h^{-1}g$ acts as the identity on M. To remedy this situation, we simply replace G by the quotient G/H where H is the subgroup of all elements of the form h for which $hx=x$ for all x. This is just a technical detail, and in most examples we will use this will not occur (for example, for $G=O(n)$, the set of $n\times n$ orthogonal matrices, and $M=S^{n-1}$, the unit sphere)

Since G acts on M by isometries, G acts on itself by isometries with right multiplication, since

$$\rho(gk,hk)=\sup_{x\in M} d(gkx,hkx)=\sup_{y=kx\in M} d(gy,hy)=\rho(g,h)$$

Also, (G,ρ) is a compact metric space since M is compact:

> Can show through diagonalization that a sequence g_n has a subsequence for which $g_{n'}x$ converges for all x. Transitivity shows that there is some g for which the limit above is gx (somehow). Since g_n are isometries, we can prove that $\rho(g_n,g)\to 0$ (look at successive ε-nets).

Now by what we just proved, there exists a measure ν on G which is G-invariant (under right-multiplication). In other words, for any $f\in C(G)$,

$$\int f(g)\,d\nu(g)=\int f(gh)\,d\nu(g)$$

Now consider any measure μ which is G-invariant, and take $f\in C(M)$. We have the following computation:

$$\begin{aligned}\nu(G)\int_M f d\mu &= \int_G\int_M f(x)\,d\mu(x)\,d\nu(g)\\ (G\text{-invariance of }\mu) &= \int_G\int_M f(gx)\,d\mu(x)\,d\nu(g)\\ (\text{Fubini}) &= \int_M\left[\int_G f(gx)\,d\nu(g)\right]d\mu(x)\end{aligned}$$

We show that the bracketed term is independent of x. Take some other $y\in M$, then by transitivity there is some $h\in G$ for which $hx=y$, and thus

$$\int_G f(gy)\,d\nu(g)=\int_G f(ghx)\,d\nu(g)=\int_G f(gx)\,d\nu(g)$$

noting that $f(\cdot\, x)$ is a continuous function on G. This means if we denote the bracketed term by $\bar{\nu}(f)$, we have that

$$\nu(G)\int_M f d\mu=\bar{\nu}(f)\int_M d\mu(x)$$

or

$$\int_M f d\mu=\frac{\bar{\nu}(f)}{\nu(G)}\,\mu(M)$$

and this implies the result (integral is determined up to the multiplicative constant $\mu(M)$).

□

Prékopa-Leindler Inequality

Theorem 5. *Let $f, g, m\colon \mathbb{R}^n \to [0, \infty)$ be integrable functions, and take $\lambda \in (0, 1)$. Assume that for all x, $y \in \mathbb{R}^n$ we have the following inequality:*

$$m(\lambda x + (1 - \lambda) y) \geq f(x)^\lambda g(y)^{1-\lambda}$$

Then

$$\int_{\mathbb{R}^n} m(x) dx \geq \left(\int_{\mathbb{R}^n} f(x) dx \right)^\lambda \left(\int_{\mathbb{R}^n} g(x) dx \right)^{1-\lambda}$$

Proof. Let μ_n denote the n dimensional Lebesgue measure. We will prove the result by induction on n. For the case $n = 1$, we use the fact that if $A, B \subset \mathbb{R}$ are nonempty and measurable, then

$$\mu_1(A + B) \geq \mu_1(A) + \mu_1(B)$$

To prove this, it suffices to prove this for compact sets by set approximation results. Also, we are allowed to translate A, B as we please, since μ_1 is translation invariant. Thus, move A so that the max value is 0, and move B so that the min value is 0, then $A \cap B = \{0\}$. Now note that since $0 \in A \cap B$, $A + B \supset A \cup B$, and this gives the result since

$$\mu_1(A + B) \geq \mu_1(A \cup B) = \mu_1(A) + \mu_1(B)$$

as A, B overlap on a set of measure zero $\{0\}$.

Continuing the proof, by approximation results again it suffices to prove the result for bounded f, g. By scaling we may assume that $\|f\|_\infty = \|g\|_\infty = 1$. Set some $t \in (0, 1)$. Let $A = \{x\colon f(x) \geq t\}$, $B = \{x\colon g(x) \geq t\}$, and $C = \{x\colon m(x) \geq t\}$. Now if $x \in A$, $y \in B$, then

$$\begin{aligned} m(\lambda x + (1 - \lambda) y) &\geq f(x)^\lambda g(y)^{1-\lambda} \\ &\geq t^\lambda t^{1-\lambda} = t \end{aligned}$$

so that $\lambda A + (1 - \lambda) B \subset C$. Note that by assumption and since $t < 1$, A, B are nonempty. Thus applying the earlier result, we have that

$$\begin{aligned} \mu_1(C) &\geq \mu_1(\lambda A + (1-\lambda) B) \\ &\geq \mu_1(\lambda A) + \mu_1((1 - \lambda) B) \\ &= \lambda \mu_1(A) + (1 - \lambda) \mu_1(B) \\ \mu_1(m \geq t) &\geq \lambda \mu_1(f \geq t) + (1 - \lambda) \mu_1(g \geq t) \end{aligned}$$

Integrating over t and using Fubini, we get that

$$\begin{aligned} \int_0^\infty \mu_1(m \geq t) dt &\geq \lambda \int_0^\infty \mu_1(f \geq t) dt + (1 - \lambda) \int_0^\infty \mu_1(g \geq t) dt \\ \int_{\mathbb{R}} m(x) dx &\geq \lambda \int_{\mathbb{R}} f(x) dx + (1 - \lambda) \int_{\mathbb{R}} g(x) dx \\ \text{(AM-GM inequality)} &\geq \left(\int_{\mathbb{R}} f(x) dx \right)^\lambda \left(\int_{\mathbb{R}} g(x) dx \right)^{1-\lambda} \end{aligned}$$

Note that for $n=1$ a stronger results holds, that we can bound by the arithmetic mean. This does not hold for larger dimensions.

For the induction step, assume the result holds for $n-1$. We will use $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$. For $(t, s) \in \mathbb{R} \times \mathbb{R}^{n-1}$, define $f_t(s) = f(t, s)$, $g_t(s) = g(t, s)$ and $m_t(s) = m(t, s)$. Take $t_0, t_1 \in \mathbb{R}$ and define $t = \lambda t_0 + (1 - \lambda) t_1$. If $x, y \in \mathbb{R}^{n-1}$ then

$$\begin{aligned} m_t(\lambda x + (1-\lambda) y) &= m(\lambda (t, x) + (1-\lambda)(t, y)) \\ &\geq f(t_0, x)^{\lambda} g(t, y)^{1-\lambda} \\ &\geq f_{t_0}(x)^{\lambda} g_{t_1}(y)^{1-\lambda} \end{aligned}$$

and thus the triple $f_{t_0}, g_{t_1},$ and m_t satisfy the assumptions for $n-1$. Induction gives

$$\int_{\mathbb{R}^{n-1}} m_t(x) \, d\mu_{n-1}(x) \geq \left(\int_{\mathbb{R}^{n-1}} f_{t_0}(x) \, d\mu_{n-1}(x) \right)^{\lambda} \left(\int_{\mathbb{R}^{n-1}} g_{t_1}(x) \, d\mu_{n-1}(x) \right)^{1-\lambda}$$

if we denote

$$\begin{aligned} \bar{m}(t) &= \int_{\mathbb{R}^{n-1}} m_t(x) \, d\mu_{n-1}(x) \\ \bar{f}(t) &= \int_{\mathbb{R}^{n-1}} f_t(x) \, d\mu_{n-1}(x) \\ \bar{g}(t) &= \int_{\mathbb{R}^{n-1}} g_t(x) \, d\mu_{n-1}(x) \end{aligned}$$

then the above says that

$$\bar{m}(\lambda t_0 + (1-\lambda) t_1) \geq \bar{f}(t_0)^{\lambda} \bar{g}(t_1)^{1-\lambda}$$

Thus the triple $\bar{m}$, $\bar{f}$, and $\bar{g}$ satisfy the hypothesis for the case $n=1$ and we have that

$$\int_{\mathbb{R}} \bar{m}(t) \, d\mu_1(t) \geq \left(\int_{\mathbb{R}} \bar{f}(t) \, d\mu_1(t) \right)^{\lambda} \left(\int_{\mathbb{R}} \bar{g}(t) \, d\mu_1(t) \right)^{1-\lambda}$$

and this is exactly the result we want. $\square$

Brunn-Minkowski Inequality

Corollary 6. (Brunn-Minkowski inequality) *Let $A, B \subset \mathbb{R}^n$ be measurable and non-empty. Then*

1. *For all $\lambda \in [0, 1]$,*

$$\operatorname{vol}(\lambda A + (1-\lambda) B) \geq \operatorname{vol}(A)^{\lambda} \operatorname{vol}(B)^{1-\lambda}$$

2. $\operatorname{vol}(A+B)^{1/n} \geq \operatorname{vol}(A)^{1/n} + \operatorname{vol}(B)^{1/n}$

Proof. For (1), we apply Prékopa-Leindler for $f = \mathbf{1}_A, g = \mathbf{1}_B, m = \mathbf{1}_{\lambda A + (1-\lambda) B}$ so that

$$m(\lambda x + (1-\lambda) y) \geq f(x)^{\lambda} g(y)^{1-\lambda}$$

since if $x \in A$ and $y \in B$ then $\lambda x + (1-\lambda)y \in \lambda A + (1-\lambda)B$, and otherwise the inequality holds trivially. Then the inequality gives

$$\mathrm{vol}(\lambda A + (1-\lambda)B) = \int_{\mathbb{R}} \mathbf{1}_{\lambda A + (1-\lambda)B}\, dx \geq \left(\int_{\mathbb{R}} \mathbf{1}_A\, dx\right)^{\lambda} \left(\int_{\mathbb{R}} \mathbf{1}_B\, dx\right)^{1-\lambda} = \mathrm{vol}(A)^{\lambda}\, \mathrm{vol}(B)^{1-\lambda}$$

(2) is a consequence of 1. First we normalize the sets so that $A = \frac{\tilde{A}}{(\mathrm{vol}(A))^{1/n}}$, $B = \frac{\tilde{B}}{(\mathrm{vol}(B))^{1/n}}$. Then we see that

$$\mathrm{vol}\left(\frac{\mathrm{vol}(A)^{1/n}}{\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}} \tilde{A} + \frac{\mathrm{vol}(B)^{1/n}}{\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}} \tilde{B}\right) \geq 1$$

(note $\mathrm{vol}(\tilde{A}) = \mathrm{vol}(\tilde{B}) = 1$). But then

$$\frac{\mathrm{vol}(A)^{1/n}}{\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}} \tilde{A} + \frac{\mathrm{vol}(B)^{1/n}}{\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}} \tilde{B} = \frac{A+B}{\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}}$$

and thus we have

$$\frac{\mathrm{vol}(A+B)}{(\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n})^n} \geq 1$$

and rearranging gives the desired result. □

Week 2 (9/14/2010)

The goal for this part of the course is Dvoretzky's Theorem:

Theorem 7. *For every $\varepsilon \in (0,1)$, there exists a constant $C(\varepsilon)$ such that the following statement holds:*

If $K \subset \mathbb{R}^n$ is a centrally symmetric convex body (compact, with nonempty interior), then there exists a linear subspace $V \subset \mathbb{R}^n$ such that

1. *Exists $r > 0$ such that*

$$r(B_2^n \cap V) \subset (K \cap V) \subset (1+\varepsilon) r (B_2^n \cap V)$$

2. $\dim V \geq c(\varepsilon) \log n$

$B_2^n = \{x \in \mathbb{R}^n \colon \sum x_i^2 \leq 1\}$ *is the Euclidean unit ball.*

Last time we covered the Brunn-Minkowski inequality (Corollary 6). There are many nontrivial consequences of this inequality.

Isoperimetric Theorem

Let $A \subset \mathbb{R}^n$, a subset with smooth boundary, and define

$$\mathrm{vol}_{n-1}(\partial A) = \lim_{\varepsilon \to 0} \frac{\mathrm{vol}(A + \varepsilon B_2^n) - \mathrm{vol}(A)}{\varepsilon}$$

Theorem 8. *If B is a Euclidean ball with the same volume as A, then*

$$\mathrm{vol}_{n-1}(\partial B) \leq \mathrm{vol}_{n-1}(\partial A)$$

We will use the notation $A_\varepsilon := A + \varepsilon B_2^n = \{x \in \mathbb{R}^n : d(x, A) \leq \varepsilon\}$.

Proof. We will show that $\mathrm{vol}(A_\varepsilon) \geq \mathrm{vol}(B_\varepsilon)$. First, let's compute the radius of $B = r B_2^n$. We have that

$$\mathrm{vol}(B) = r^n \,\mathrm{vol}(B_2^n) = \mathrm{vol}(A)$$

and thus

$$r = \left[\frac{\mathrm{vol}(B)}{\mathrm{vol}(A)}\right]^{1/n}$$

Now using Brunn-Minkowski,

$$\begin{aligned} \mathrm{vol}(A_\varepsilon)^{1/n} &= \mathrm{vol}(A + \varepsilon B_2^n)^{1/n} \\ (\text{Brunn-Minkowski}) &\geq \mathrm{vol}(A)^{1/n} + \mathrm{vol}(\varepsilon B_2^n)^{1/n} \\ &= r\,\mathrm{vol}(B_2^n)^{1/n} + \varepsilon\,\mathrm{vol}(B_2^n)^{1/n} \\ &= \mathrm{vol}((r+\varepsilon)B_2^n)^{1/n} \\ &= \mathrm{vol}(B_\varepsilon)^{1/n} \end{aligned}$$

This shows that

$$\frac{\mathrm{vol}(A_\varepsilon) - \mathrm{vol}(A)}{\varepsilon} \geq \frac{\mathrm{vol}(B_\varepsilon) - \mathrm{vol}(B)}{\varepsilon}$$

and thus $\mathrm{vol}_{n-1}(\partial B) \leq \mathrm{vol}_{n-1}(\partial A)$ (taking limit as $\varepsilon \to 0$).

□

Kahane's Inequality

Theorem 9. $\frac{1}{2}\,\mathbb{E}_\varepsilon\big\|\sum_{i=1}^n \varepsilon_i x_i\big\|\frac{1}{2}\,\mathbb{E}_\varepsilon\big\|\sum_{i=1}^n \varepsilon_i x_i\big\|$ *For all $\infty > p \geq 1$, there exists $K_p > 0$ such that the following statement holds:*

If $(X, \|\cdot\|)$ is any normed space and $x_1, \ldots, x_n \in X$, then

$$\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\| \leq \left(\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|^p\right)^{1/p} \leq K_p\,\mathbb{E}\left\|\sum_{i=1}^n \varepsilon_i x_i\right\|$$

$\mathbb{E}$ is expectation with respect to a random variable $\varepsilon \in \{\pm 1\}^n$.

Remark: It can be shown that $K_p \approx \sqrt{p}$, but we will give a proof with $K_p \approx p$. Also, the first inequality follows directly from convexity (Jensen).

Definition 10. *If μ is a Borel measure on $\mathbb{R}^n$ we say that μ is log-concave if for all A, B measurable sets, we have*

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}$$

for all $\lambda \in (0,1)$

Example 11. If $L \subset \mathbb{R}^n$ is a convex body, let μ be the normalized volume on L:

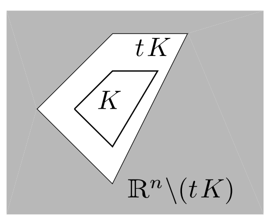
$$\mu(A) = \frac{\mathrm{vol}(A \cap L)}{\mathrm{vol}(L)}$$

μ is a log-concave probability measure on $\mathbb{R}$:

$$\begin{aligned}
(\lambda A + (1-\lambda)B) \cap L &\supset \lambda(A \cap L) + (1-\lambda)(B \cap L) \\
\frac{\mathrm{vol}((\lambda A + (1-\lambda)B) \cap L)}{\mathrm{vol}(L)} &\geq \frac{\mathrm{vol}(\lambda(A \cap L) + (1-\lambda)(B \cap L))}{\mathrm{vol}(L)} \\
&\geq \left(\frac{\mathrm{vol}(A \cap L)}{\mathrm{vol}(L)}\right)^{\lambda} \left(\frac{\mathrm{vol}(B \cap L)}{\mathrm{vol}(L)}\right)^{1-\lambda}
\end{aligned}$$

Lemma 12. (Borell's Inequality) *Let K be a centrally symmetric convex body, μ a log-concave probability measure on $\mathbb{R}^n$, $t > 1$. Then*

$$\mu(\mathbb{R}^n \backslash (tK)) \leq \mu(K) \left(\frac{1}{\mu(K)} - 1\right)^{\frac{t+1}{2}}$$



Proof. We claim that

$$\mathbb{R}^n \backslash K \supset \frac{2}{t+1} (\mathbb{R}^n \backslash (tK)) + \frac{t-1}{t+1} K$$

Suppose not, then there exists $a \in K$, $b \notin tK$ such that

$$\frac{2}{t+1} b + \frac{t-1}{t+1} a = c \in K$$

This means that

$$b = \frac{t+1}{2} c - \frac{t-1}{2} a$$

$$\frac{1}{t} b = \frac{t+1}{2t} c + \frac{t-1}{2t} (-a)$$

with $t > 1$. Since $a \in K$, $-a \in K$ by symmetry, and $c \in K$ already. Thus by convexity, $\frac{1}{t} b \in K$, which contradicts $b \notin tK$.

Now we can apply Brunn-Minkowski:

$$\begin{aligned} 1-\mu(K)=\mu(\mathbb{R}^n\backslash K) &\geq \mu\left(\frac{2}{t+1}(\mathbb{R}^n\backslash(tK))+\frac{t-1}{t+1}K\right)\\ &\geq \mu(\mathbb{R}^n\backslash(tK))^{\frac{2}{t+1}}\mu(K)^{\frac{t-1}{t+1}}\\ &\geq \mu(\mathbb{R}^n\backslash(tK))^{\frac{2}{t+1}}\mu(K) \end{aligned}$$

and this implies that

$$\mu(\mathbb{R}^n\backslash(tK))\leq\left[\frac{1}{\mu(K)}-1\right]^{\frac{t+1}{2}}$$

□

Now let $L=\left[-\frac{1}{2},\frac{1}{2}\right]^n$, and $\mu(A)=\frac{\mu(A\cap L)}{\mu(L)}$, $(X,\|\cdot\|)$ a normed space, and $x_1,\ldots,x_n\in X,\ \ a_1,\ldots,a_n\in L$. Assume that

$$\int_L\left\|\sum_{i=1}^n a_i x_i\right\|da=1$$

Define $K=\left\{a\in\mathbb{R}^n\colon\ \left\|\sum_{i=1}^n a_i\ x_i\right\|\leq 3\right\}$. K is convex and centrally symmetric. Then by Borell's inequality, we have that

$$\begin{aligned} \mu\left(a\in L\colon\left\|\sum_{i=1}^n a_i x_i\right\|>3t\right) &\leq \mu(K)\left(\frac{1}{\mu(K)}-1\right)^{\frac{t+1}{2}}\\ &\leq 1\cdot\left(\frac{1}{2/3}-1\right)^{\frac{t+1}{2}}\\ &= \left(\frac{1}{2}\right)^{\frac{t+1}{2}} \end{aligned}$$

where we note that $\mu(K)\geq\frac{2}{3}$ by Markov's inequality, using the assumption:

$$\mu(K^c)=\mu\left\{\left\|\sum_{i=1}^n a_i x_i\right\|>3\right\}\leq\frac{\int_L\left\|\sum_{i=1}^n a_i x_i\right\|da}{3}=\frac{1}{3}$$

Now we integrate:

$$\begin{aligned} \int_{\mathbb{R}^n}\left\|\sum_{i=1}^n a_i x_i\right\|^p d\mu(a) &= \int_0^\infty pt^{p-1}\mu\left(\left\|\sum_{i=1}^n a_i x_i\right\|>t\right)dt\\ &\leq \int_0^3 pt^{p-1}dt+\int_3^\infty pt^{p-1}\left(\frac{1}{2}\right)^{\frac{t/3+1}{2}}dt\\ &\leq K_p \end{aligned}$$

The first step holds by Fubini, and for the rest, the important part is that towards ∞ we have exponential decay, and thus it is integrable. Thus (after scaling) we have proved:

$$\left(\int_{\mathbb{R}^n}\left\|\sum_{i=1}^n a_i x_i\right\|^p d\mu(a)\right)^{1/p}\leq K_p\int_L\left\|\sum_{i=1}^n a_i x_i\right\|da \qquad (*)$$

Note that the distribution of μ is that a_i is independent and uniform on $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

To transfer this result to ± 1 random variables as in the statement of Kahane's inequality (Theorem 9), we need the following useful trick:

Lemma 13. (Contraction Principle) *Let $(X, \|\cdot\|)$ be a normed space, $a_1, \ldots, a_n \in \mathbb{R}^n$, and $x_1, \ldots, x_n \in X$ with $p \geq 1$. Then*

$$\left(\mathbb{E}_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i a_i x_i \right\|^p\right)^{1/p} \leq \left(\max_{1 \leq i \leq n} |a_i|\right) \left(\mathbb{E}_\varepsilon \left\| \sum_{i=1}^n \varepsilon_i x_i \right\|^p\right)^{1/p}$$

where ε is the uniform random variable on $\{\pm 1\}^n$.

Proof. First, normalize so that $|a_i| \leq 1$ for all i (the general result follows by scaling). We will prove this by induction on the number of i such that $|a_i| \neq 1$.

If $a_i = \pm 1$ for all i, then we note that $\sum a_i \varepsilon_i x_i$ has the same distribution as $\sum \varepsilon_i x_i$, and we have equality above.

Otherwise, take $|a_1| < 1$, and we write $a_1 = \lambda \cdot 1 + (1 - \lambda) \cdot (-1)$. Then we have that

$$\sum_{i=1}^n \varepsilon_i a_i x_i = \lambda \left(\varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i\right) + (1 - \lambda) \left(-\varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i\right)$$

Convexity of $\|\cdot\|$ gives

$$\left\| \sum_{i=1}^n \varepsilon_i a_i x_i \right\| \leq \lambda \left\| \varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i \right\| + (1 - \lambda) \left\| -\varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i \right\|$$

Taking p norms in ε:

$$\begin{aligned}
\mathbb{E}_\varepsilon \left[\left\| \sum_{i=1}^n \varepsilon_i a_i x_i \right\|^p\right]^{1/p} &\leq \mathbb{E}_\varepsilon \left[\left(\lambda \left\| \varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i \right\| + (1 - \lambda) \left\| -\varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i \right\|\right)^p\right]^{1/p} \\
&\leq \lambda \mathbb{E}_\varepsilon \left[\left\| \varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i \right\|^p\right]^{1/p} + (1 - \lambda) \mathbb{E}_\varepsilon \left[\left\| -\varepsilon_1 x_1 + \sum_{i \geq 2} \varepsilon_i a_i x_i \right\|^p\right]^{1/p} \\
\text{(induction)} &\leq \lambda \cdot |1| \cdot \mathbb{E}_\varepsilon \left[\left\| \sum_{i=1}^n \varepsilon_i x_i \right\|^p\right]^{1/p} + (1 - \lambda) \cdot |-1| \cdot \mathbb{E}_\varepsilon \left[\left\| \sum_{i=1}^n \varepsilon_i x_i \right\|^p\right]^{1/p} \\
&= \mathbb{E}_\varepsilon \left[\left\| \sum_{i=1}^n \varepsilon_i x_i \right\|^p\right]^{1/p}
\end{aligned}$$

where we have used convexity of $\mathbb{E}_\varepsilon[(\cdot)^p]^{1/p}$ and induction.

$\square$

Now we can transfer between a_i uniform on $L = \left[-\frac{1}{2}, \frac{1}{2}\right]^n$ and ε_i uniform on $\{\pm 1\}^n$. First note that

$$\int_{\mathbb{R}^n} \left\| \sum_{i=1}^n a_i x_i \right\| d\mu(a) = \int_{\mathbb{R}^n} \mathbb{E}_\varepsilon \left\| \sum_{i=1}^n a_i \varepsilon_i x_i \right\| d\mu(a)$$

since the distribution of a_i is equal to the distribution of $a_i\,\varepsilon_i$ (in particular, the equation above is true without the $\mathbb{E}_\varepsilon$). Then applying the contraction principle, we have that

$$\int_{\mathbb{R}^n} \mathbb{E}_\varepsilon\left\|\sum_{i=1}^n a_i\,\varepsilon_i\,x_i\right\| d\mu(a) \le \frac{1}{2}\,\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^n \varepsilon_i\,x_i\right\|\right]\int_{\mathbb{R}^n} d\mu(a) = \frac{1}{2}\,\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n \varepsilon_i\,x_i\right\|$$

Along with $(*)$, this gives the second inequality of Theorem 9.

To get the other inequality, we note that

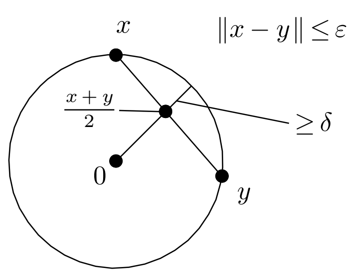
$$\begin{aligned}
\left(\int_{\mathbb{R}^n}\left\|\sum_{i=1}^n a_i\,x_i\right\|^p d\mu(a)\right)^{1/p} &= \left(\int_{\mathbb{R}^n}\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^n |a_i|\,\varepsilon_i\,x_i\right\|^p\right] d\mu(a)\right)^{1/p} \\
(\text{Minkowski}) \quad &\ge \left(\mathbb{E}_\varepsilon\left\|\int_{\mathbb{R}^n}\sum_{i=1}^n |a_i|\,\varepsilon_i\,x_i\,d\mu(a)\right\|^p\right)^{1/p} \\
&= \left(\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n\left(\int_{-1/2}^{1/2}|a_i|\,da_i\right)\varepsilon_i\,x_i\right\|^p\right)^{1/p} \\
&= \left(\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n \frac{1}{4}\varepsilon_i\,x_i\right\|^p\right)^{1/p} \\
(\text{Jensen}) \quad &\ge \frac{1}{4}\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n \varepsilon_i\,x_i\right\|
\end{aligned}$$

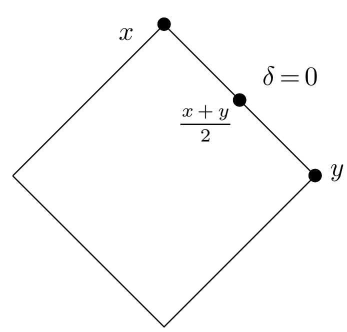
Concentration of Measure on the Sphere

Let $(X, \|\cdot\|)$ be a normed space, and $K = \{x \in X\colon \|x\| \le 1\}$.

Definition 14. *The modulus of uniform convexity for the normed space* $(X, \|\cdot\|)$ *is*

$$\delta_{\|\cdot\|}(\varepsilon) = \inf\left\{1 - \left\|\frac{x+y}{2}\right\| : \|x\|, \|y\| \le 1,\ \|x-y\| \le \varepsilon\right\}$$





The left picture shows an example where $\delta > 0$ (Euclidean norm), and the right shows the kind of spaces we want to avoid (l^1 norm). When $\delta = 0$, there are "flat edges".

Example 15. If X is a Hilbert space, then we have the parallelogram identity:

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

and if $\|x\| = \|y\| = 1$, and $\|x - y\| \geq \varepsilon$, then

$$\|x + y\|^2 + \varepsilon^2 \leq 4$$

so that $\left\|\frac{x+y}{2}\right\| \leq \sqrt{1 - \frac{\varepsilon^2}{4}} \approx 1 - \frac{\varepsilon^2}{8}$.

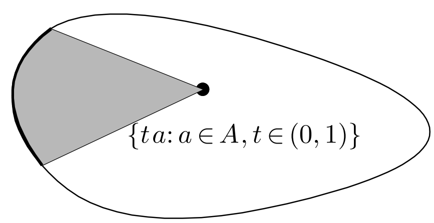
It is a nontrivial result to show that if for a normed space, $\delta(\varepsilon) > \frac{\varepsilon^2}{8}$, then it must be a Hilbert space. In other words, Hilbert spaces have the best modulus of uniform convexity.

Now let K be a centrally symmetric convex body, and let $\partial K = S$.

Define

$$\begin{aligned}
\nu(A) &= \frac{\mathrm{vol}(A \cap K)}{\mathrm{vol}(K)}, A \subset K \\
\mu(A) &= \frac{\mathrm{vol}(\{ta : a \in A, t \in (0,1)\}}{\mathrm{vol}(K)}, A \subset S
\end{aligned}$$

μ is the "cone measure" on S:



Fact: If $K = B_2^n$, then μ is the surface measure (normalized) on $\mathbb{S}^{n-1}$. This can be checked directly, but one way to see this quickly is to use the uniqueness of the Haar measure. Since $O(n)$ the group of orthogonal $n \times n$ matrices acts transitively on $\mathbb{S}^{n-1}$, and μ is invariant under $O(n)$, μ must be the surface measure on $\mathbb{S}^{n-1}$ (which is also invariant under $O(n)$).

Exercise 1. If $K = B_1^n$ or $K = B_\infty^n$ then μ is the surface area measure. This is not true for $p \notin \{1, 2, \infty\}$.

Theorem 16. (Gromov-Milman Theorem) *Let $K \subset \mathbb{R}^n$ be centrally symmetric convex body, and $S = \partial K$. Let $\delta = \delta_{\|\cdot\|}$. Then*

1. *For all $A \subset K$, $\nu(A_\varepsilon) \geq 1 - \frac{1}{\nu(A)} e^{-2n\delta(\varepsilon)}$*

2. *For all $A \subset S$, $\mu(A_\varepsilon) \geq 1 - \frac{2}{\mu(A)} e^{-2n\delta(\varepsilon/4)}$*

Here $A_\varepsilon = A + \varepsilon K = \{x \in \mathbb{R}^n : d(x, A) \leq \varepsilon\}$ and $d(x, y) = \|x - y\|$.

If $K = B_2^n$, then $S = \mathbb{S}^{n-1}$. Then if $\mu(A) = \frac{1}{2}$ then $\mu(A_\varepsilon) \geq 1 - ce^{-c' n \varepsilon^2}$. This is a highly unintuitive fact! For very large n, this result says that if you start with a set with half the measure, and increase by a small amount, you end up with almost everything.

Proof. (Ball, de-Renga, Villa)

(1) Let $A \subset K$. Define $B = K \backslash A_\varepsilon$. If $a \in A$ and $b \in B$ then $\|a - b\| \geq \varepsilon$, and

$$\left\| \frac{a+b}{2} \right\| \leq 1 - \delta(\varepsilon)$$

This shows that

$$\frac{A+B}{2} \subset (1 - \delta(\varepsilon))K$$

and by Brunn-Minkowski,

$$\nu(A)^{1/2}\nu(B)^{1/2} \leq \nu\left(\frac{A+B}{2}\right) \leq \nu((1-\delta(\varepsilon))K) = (1-\delta(\varepsilon))^n \leq e^{-n\delta(\varepsilon)}$$

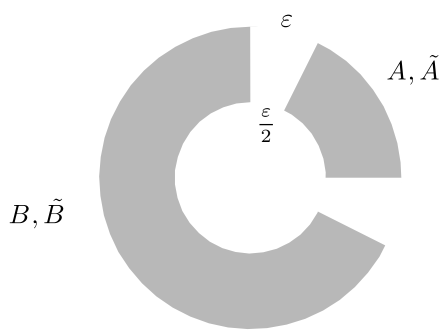
This implies that

$$1 - \nu(A_\varepsilon) = \nu(B) \leq \frac{1}{\nu(A)} e^{-2n\delta(\varepsilon)}$$

(2) Now let's take $A \subset S$, and again use $B = S \backslash A_\varepsilon$. Define the partial cones

$$\tilde{A} = \left\{ ta : a \in A, \frac{1}{2} \leq t \leq 1 \right\}, \ \tilde{B} = \left\{ tb : b \in B, \frac{1}{2} \leq t \leq 1 \right\}$$

If $\tilde{a} \in \tilde{A}$ and $\tilde{b} \in \tilde{B}$, then $\|\tilde{a} - \tilde{b}\| \geq \frac{\varepsilon}{4}$. We will show this algebraically, but it is intuitively true from examining the picture:



The smallest gap is $\varepsilon/2$ in this picture of the case when K is the circle. Now let $\tilde{a} = \alpha a$ and $\tilde{b} = \beta b$, so that $\alpha, \beta \in [1/2, 1]$ and $a \in A$, $b \in B$. We split this into cases:

- If $|\alpha - \beta| \geq \frac{\varepsilon}{4}$, then we are done by the triangle inequality

$$\|\tilde{a} - \tilde{b}\| \geq |\alpha - \beta| \geq \frac{\varepsilon}{4}$$

- Otherwise, $|\alpha - \beta| \leq \frac{\varepsilon}{4}$, and we have

$$\begin{aligned} \|\tilde{a} - \tilde{b}\| &= \|\alpha(a-b) + (\alpha - \beta)b\| \\ &\geq \|\alpha(a-b)\| - \|(\alpha-\beta)b\| \\ &\geq \frac{1}{2}\varepsilon - \frac{\varepsilon}{4} = \frac{\varepsilon}{4} \end{aligned}$$

This implies that $\frac{\tilde{A}+\tilde{B}}{2} \subset \left(1 - \delta\left(\frac{\varepsilon}{4}\right)\right)K$, and with the same computation as before, we have that

$$\nu(\tilde{B}) \leq \frac{1}{\nu(\tilde{A})} e^{-2n\delta(\varepsilon/4)}$$

To finish the result, we note that

$$\begin{aligned}
\mu(A) &= \nu([0,1]A)\\
\nu(\tilde{A}) &= \nu([0,1]A\backslash[0,1/2]A)\\
&= \mu(A)-\nu\left(\frac{1}{2}[0,1]\,A\right)\\
&= \left(1-\frac{1}{2^n}\right)\mu(A)
\end{aligned}$$

and this means that

$$\left(1-\frac{1}{2^n}\right)^2\mu(B)\leq\frac{1}{\mu(A)}\,e^{-2n\delta(\varepsilon/4)}$$

Since $B=S\backslash A_{\varepsilon}$, we have that

$$\mu(A_{\varepsilon})\geq 1-\frac{1}{\mu(A)}\,e^{-2n\delta(\varepsilon/4)}\left(1-\frac{1}{2^n}\right)^{-2}\approx 1-\frac{1}{\mu(A)}\,e^{-2n\delta(\varepsilon/4)}$$

□

Now with μ being the normalized Haar measure on $\mathbb{S}^{n-1}$, we have the following corollary:

Corollary 17. *Let $f\colon\mathbb{S}^{n-1}\rightarrow\mathbb{R}$ be an L-Lipschitz map, i.e. for all $x,y\in\mathbb{S}^{n-1}$ we have that*

$$|f(x)-f(y)|\leq L\,\|x-y\|_2$$

Let M be a median of f, i.e. $\mu(f\geq M)$, $\mu(f\leq M)\geq\frac{1}{2}$ (which always exists for any probability measure). Then for all $\varepsilon>0$,

$$\mu(x\in\mathbb{S}^{n-1}\colon|f(x)-M|\geq\varepsilon)\leq Ce^{-c'n\varepsilon^2/L^2}$$

This says that a Lipschitz function is essentially a constant on $\mathbb{S}^{n-1}$ for large n.

Proof. Let $A=\{x\in\mathbb{S}^{n-1}\colon f(x)\leq M\}$. Then $x\in A_{\varepsilon/L}$ if and only if there exists $y\in A$ with $\|x-y\|\leq\varepsilon/L$, in which case $|f(x)-f(y)|\leq\varepsilon$ and $f(x)\leq M+\varepsilon$. This means that

$$\{x\in\mathbb{S}^{n-1}\colon f(x)>M+\varepsilon\}\subset\mathbb{S}^{n-1}\backslash A_{\varepsilon/L}$$

Since $\mu(A)\geq\frac{1}{2}$, we have that

$$\mu(x\in\mathbb{S}^{n-1}\colon f(x)-M>\varepsilon)\leq Ce^{-c'n(\varepsilon/L)^2}$$

applying the previous result. By symmetry, examining $B=\{x\in\mathbb{S}^{n-1}\colon f(x)\geq M\}$, we have the opposite inequality, that

$$\mu(x\in\mathbb{S}^{n-1}\colon f(x)-M<-\varepsilon)\leq Ce^{-c'n(\varepsilon/L)^2}$$

and since the union of these sets is $\mu(x\in\mathbb{S}^{n-1}\colon\|f(x)-M\|>\varepsilon)$ the result follows by subadditivity.

□

The same result holds for the mean as well:

Corollary 18. *If $f:\mathbb{S}^{n-1}\to\mathbb{R}$, and is L-Lipschitz,*

$$\mu\left(x\in\mathbb{S}^{n-1}:\left|f(x)-\int f d\mu\right|\geq\varepsilon\right)\leq Ce^{-c' n\varepsilon^2/L^2}$$

We will show this next time.

Week 3 (9/21/2010)

Proof. (of Corollary 18) Without loss of generality, $L=1$ (by scaling). Denote

$$E:=\int_{\mathbb{S}^{n-1}} f d\mu$$

Take the product measure $\mu\times\mu$ on $\mathbb{S}^{n-1}\times\mathbb{S}^{n-1}$. Note

$$\begin{aligned}(\mu\times\mu)\big((x,y)\in\mathbb{S}^{n-1}\times\mathbb{S}^{n-1}: |f(x)-f(y)|\geq\varepsilon\big) &\leq (\mu\times\mu)\left(\left\{|f(x)-M|\geq\frac{\varepsilon}{2}\right\}\cup\left\{|f(y)-M|\geq\frac{\varepsilon}{2}\right\}\right)\\ &\leq 2\mu\left(\left\{|f(x)-M|\geq\frac{\varepsilon}{2}\right\}\right)\\ &\leq 2ae^{-bn\varepsilon^2/4}\end{aligned}$$

This means, for most pair of points, $f(x)$ and $f(y)$ are close. Now fix $\lambda>0$, we will be using a Chernoff bound:

$$\mu(|f(x)-E|>\varepsilon)=\mu\left(e^{\lambda^2|f(x)-E|^2}\geq e^{\lambda^2\varepsilon^2}\right)\leq e^{-\lambda^2\varepsilon^2}\int_{\mathbb{S}^{n-1}} e^{\lambda^2|f(x)-E|^2}d\mu(x)$$

Note that since $e^{\lambda^2t^2}$ is convex, Jensen gives

$$e^{\lambda^2|f(x)-\int f(y)d\mu(y)|^2}\leq\int_{\mathbb{S}^{n-1}} e^{\lambda^2|f(x)-f(y)|^2}d\mu(y)$$

and hence

$$\int_{\mathbb{S}^{n-1}} e^{\lambda^2|f(x)-E|^2}d\mu(x)\leq\int_{\mathbb{S}^{n-1}}\int_{\mathbb{S}^{n-1}} e^{\lambda^2|f(x)-f(y)|^2}d\mu(x)d\mu(y)$$

Now using Fubini, note that if ψ is an increasing function from $\mathbb{R}^+\to\mathbb{R}^+$, then

$$\mathbb{E}[\psi]=\int_0^\infty \psi'(t)P(X\geq t)dt$$

Thus

$$\begin{aligned}\int_{\mathbb{S}^{n-1}}\int_{\mathbb{S}^{n-1}} e^{\lambda^2|f(x)-f(y)|^2}d\mu(x)d\mu(y) &\leq \int_0^\infty 2\lambda^2te^{\lambda^2t}(\mu\times\mu)(\{(x,y): |f(x)-f(y)\geq t|\})dt\\ &\leq \int_0^\infty 2\lambda^2te^{\lambda^2t}2ae^{-bnt^2/4}dt\\ \left(\lambda^2=\frac{bn}{8}\right) &= 2a\int_0^\infty\frac{bn}{8}2te^{-bnt^2/8}dt\\ &= 2a\end{aligned}$$

Putting everything together, we conclude that

$$\mu(|f(x)-E|>\varepsilon)\leq 2ae^{-bn\varepsilon^2/8}$$

$\square$

Theorem 19. (Johnson-Lindenstrauss) *For all $\varepsilon \in (0,1)$, there exists $C(\varepsilon) > 0$ such that for all n, the following holds:*

If $x_1, \ldots, x_n$ are arbitrary points in a Hilbert space $\mathcal{H}$, then there exists $K \leq C(\varepsilon) \log n$ and $y_1, \ldots, y_n \in l_2^K$ such that

$$\|x_i - x_j\| \leq \|y_i - y_j\| \leq (1+\varepsilon)\|x_i - x_j\| \quad \textit{for all } i, j$$

We may assume that $x_1, \ldots, x_n \in l_2^n$ since we can restrict our attention to the span, which is finite dimensional, and norms on finite dimensional spaces are all equivalent. The conclusion is that essentially we can map n points to a significantly lower dimensional space K while roughly preserving the distances between the points.

Proof. Let ν be the normalized Haar measure on $O(n)$, the orthogonal group. Let P_0 be the orthogonal projection onto the span of $e_1, \ldots, e_K$, i.e.

$$Px = (x_1, \ldots, x_k, 0, \ldots, 0)$$

Fix $x_0 \in \mathbb{S}^{n-1}$, we will look at $U^{-1} P_0 U x_0$ for $U \in O(n)$ (a random K-dimensional projection).

Observation: The random variable $U x_0$ on $O(n)$ has the same distribution as $x \in \mathbb{S}^{n-1}(\mu)$, identifying the points $x(U) = U x_0$. In fact, for $A \subset \mathbb{S}^{n-1}$,

$$\mu(A) = \nu(U \in O(n) \colon U x_0 \in A)$$

(if we define $\tilde{\mu}$ to be the RHS above, it is $O(n)$-invariant (since ν is), and $O(n)$ is transitive, so that the $O(n)$-invariant Haar measure is unique).

Now let

$$E = \int_{\mathbb{S}^{n-1}} \|P_0 x\|_2 \, d\mu(x) = \int_{O(n)} \|P_0 U x_0\|_2 \, d\nu(U)$$

Exercise 2. Compute E directly.

Note μ is normalized, so to transfer to polar coordinates we need to throw back the surface area of $\mathbb{S}^{n-1}$, denote this by $S_{n-1} = n \frac{\pi^{n/2}}{\Gamma(n/2+1)}$. Polar coordinates with Gaussian:

$$\begin{aligned} \frac{S_{n-1}}{(2\pi)^{n/2}} \int_0^\infty r^n e^{-r^2/2} dr \int_{\mathbb{S}^{n-1}} \|P_0 x\|_2 \, d\mu(x) &= \frac{1}{(2\pi)^{n/2}} \int_0^\infty \int_{\mathbb{S}^{n-1}} \|P_0 (rx)\|_2 \, e^{-r^2/2} \, r^{n-1} dr \, S_{n-1} d\mu(x) \\ &= \int_{\mathbb{R}^n} \|P_0 x\|_2 \, d\gamma_n \end{aligned}$$

Note we can do a change of variable $u = \frac{r^2}{2}$, $du = r\,dr$, to get

$$\begin{aligned} \frac{S_{n-1}}{(2\pi)^{n/2}} \int_0^\infty r^n e^{-r^2/2} dr &= \frac{S_{n-1}}{(2\pi)^{n/2}} \int_0^\infty (2u)^{\frac{n+1}{2}-1} e^{-u} du \\ &= \frac{n}{\Gamma(n/2+1)} 2^{-1/2} \Gamma\left(\frac{n+1}{2}\right) \\ &\sim \frac{n}{\sqrt{\pi n}\,(n/2e)^{n/2}} \sqrt{\frac{4\pi}{n+1}} \left(\frac{n+1}{2e}\right)^{\frac{n+1}{2}} \\ &= C\sqrt{n} \left(\frac{2e(n+1)}{n}\right)^{n/2} \sim C\sqrt{n} \end{aligned}$$

applying Stirling's approximation.

Now

$$\begin{aligned}\int_{\mathbb{R}^n} \|P_0 x\|_2 \, d\gamma_n &= \int_{\mathbb{R}^K} \|x\|_2 \, d\gamma_K \int_{\mathbb{R}^{n-K}} d\gamma_{n-K} \\ &= \frac{S_{K-1}}{(2\pi)^{K/2}} \int_0^\infty r^K e^{-r^2/2} dr \\ &= \frac{K\,\Gamma(K/2+1/2)}{\sqrt{2}\,\Gamma(K/2+1)} \sim C\sqrt{K}\end{aligned}$$

Thus,

$$E = \frac{\Gamma(n/2+1)}{n\Gamma(n/2+1/2)} \cdot \frac{K\,\Gamma(K/2+1/2)}{\Gamma(K/2+1)} \sim C\sqrt{\frac{K}{n}}$$

We can actually get away without computing E with some concentration estimates.

Claim: $E \geq C\sqrt{K/n}$.

Computing the second moment gives:

$$\begin{aligned}\int_{\mathbb{S}^{n-1}} \|P_0 x\|_2^2 \, d\mu(x) &= \int_{\mathbb{S}^{n-1}} \left(\sum_{i=1}^K x_i^2 \right) d\mu(x) \\ &= \sum_{i=1}^K \int_{\mathbb{S}^{n-1}} x_i^2 \, d\mu(x) = K \int_{\mathbb{S}^{n-1}} x_1^2 \, d\mu(x) \\ &= \frac{K}{n} \int_{\mathbb{S}^{n-1}} \left(\sum_{i=1}^n x_i^2 \right) d\mu(x) = \frac{K}{n}\end{aligned}$$

where we have used the rotation invariance of μ to show that $\int_{\mathbb{S}^{n-1}} x_i^2 \, d\mu(x) = \int_{\mathbb{S}^{n-1}} x_1^2 \, d\mu(x)$.

We also compute the fourth moment:

$$\begin{aligned}\int_{\mathbb{S}^{n-1}} \|P_0 x\|_2^4 \, d\mu(x) &= \int_0^\infty 4u^3 \, \mu(\|P_0 x\|_2 \geq u) \, du \\ &\leq \int_0^{2E} 4u^3 \, du + \int_{2E}^\infty 4u^3 \, \mu\Big(| \|P_0 x\| - E | \geq \frac{u}{2} \Big) \, du\end{aligned}$$

since if $u > 2E$ and $\|P_0 x\| \geq u$, then $\|P_0 x\| - E \geq u - \frac{u}{2} = \frac{u}{2}$. Continuing,

$$\begin{aligned} &\lesssim E^4 + \int_{2E}^\infty u^3 e^{-bnu^2} du \\ (nu^2 = v^2) \quad &= E^4 + \frac{1}{n^2} \int_{2E\sqrt{n}}^\infty v^3 e^{-bv^2} dv \\ &\lesssim E^4 + \frac{1}{n^2} \\ &\lesssim \left(\int_{\mathbb{S}^{n-1}} \|P_0 x\|_2^2 \, d\mu(x) \right)^2\end{aligned}$$

where we used the second moment computation above and the fact that

$$E^4 = \left(\int_{\mathbb{S}^{n-1}} \|P_0 x\| \, d\mu(x) \right)^4 \leq \left(\int_{\mathbb{S}^{n-1}} \|P_0 x\|^2 \, d\mu(x) \right)^2$$

by Jensen (convexity of $|\cdot|^2$). Now letting $Z = \|P_0 x\|$ be a random variable on $\mathbb{S}^{n-1}$, i.e. $E = \mathbb{E}[Z]$, we have just shown that $\mathbb{E}[Z^4] < C(\mathbb{E}[Z^2])^2$, and this also gives an estimate for comparing the second moment with the first using Hölder with $p = 3/2, q = 3$:

$$\mathbb{E}[Z^2] = \mathbb{E}[Z^{2/3} Z^{4/3}] \leq (\mathbb{E}[Z])^{2/3} (\mathbb{E}[Z^4])^{1/3} \leq (\mathbb{E}[Z])^{2/3} C^{1/3} (\mathbb{E}[Z^2])^{2/3}$$

The last inequality above uses the estimate between the fourth and second moments. Finally,

$$\mathbb{E}[Z^2] \le C(\mathbb{E}[Z])^2$$

and therefore

$$E = \mathbb{E}[Z] \ge \frac{1}{\sqrt{C}}(\mathbb{E}[Z^2])^{1/2} \gtrsim \sqrt{\frac{K}{n}}$$

Since $x \mapsto \|P_0\, x\|_2$ is a 1-Lipschitz function, we have the following concentration result for $\|P_0\, x\|$ around its mean E:

$$\mu\big(x \in \mathbb{S}^{n-1}\colon |\, \|P_0\, x\|_2 - E\,| \ge \varepsilon E\big) \le a e^{-bn\varepsilon^2 E^2} \le a e^{-b^2 K \varepsilon^2} \quad (nE^2 \ge K)$$

This transfers by the uniqueness of Haar measure to a concentration result for the random projection $\|U^{-1} P_0\, U x_0\|_2$:

$$\nu\big(U \in O(n)\colon \big|\|U^{-1}P_0\, Ux_0\|_2 - E\big| \ge \varepsilon E\big) = \mu\big(x \in \mathbb{S}^{n-1}\colon |\, \|P_0\, x\|_2 - E\,| \ge \varepsilon E\big) \le a e^{-b^2 K \varepsilon^2}$$

(note $\|U^{-1}\, P_0\, Ux_0\|_2 = \|P_0\, Ux_0\|_2$ since U is orthogonal).

We will plug in $x_0 = \frac{x_i - x_j}{\|x_i - x_j\|_2}$ for each pair i, j, and this will give the result. The union bound tells us that at least one of the $\binom{n}{2}$ pairs x_i, x_j fails to satisfy $\left|\left\|U^{-1} P_0\, U\, \frac{x_i - x_j}{\|x_i - x_j\|_2}\right\|_2 - E\right| \le \varepsilon E$ with probability bounded above by $\binom{n}{2}\, a e^{-b^2 K \varepsilon^2}$. This probability is < 1 if $K > \frac{C}{\varepsilon^2} \log n$, in which case there must exist some $U \in O(n)$ for which all of the pairs satisy $\left|\left\|U^{-1} P_0\, U\, \frac{x_i - x_j}{\|x_i - x_j\|_2}\right\|_2 - E\right| \le \varepsilon E$.

Setting $y_i = \frac{U^{-1} P_0\, U x_i}{(1-\varepsilon) E}$, this means that

$$\begin{aligned} \left|\left\|\frac{(1-\varepsilon)E(y_i - y_j)}{\|x_i - x_j\|_2}\right\|_2 - E\right| &\le \varepsilon E \\ |(1-\varepsilon)\|y_i - y_j\|_2 - \|x_i - x_j\|_2| &\le \varepsilon \|x_i - x_j\|_2 \end{aligned}$$

or

$$\|x_i - x_j\|_2 \le \|y_i - y_j\|_2 \le \frac{1+\varepsilon}{1-\varepsilon} \|x_i - x_j\|_2$$

for all i, j, which is the desired result.

□

Remark 20. We proved the JL Lemma above with $K \approx \frac{\log n}{\varepsilon^2}$. This is almost sharp. Alon proved the result with $K \ge c \frac{\log n}{\varepsilon^2 \log(1/\varepsilon)}$. It is still accessible, but needs more work.

Now we are working towards proving Dvoretsky's Theorem (Theorem 7).

Terminology: A subset $\mathcal{N}$ in a metric space is called ε-dense if for all $x \in X$, there exists $y \in \mathcal{N}$ such that $d(x, y) < \varepsilon$. In other words, $\mathcal{N}_\varepsilon = X$.

Lemma 21. *Let X be a finite dimensional normed space, Y a normed space, and $T\colon X \to Y$ a linear operator. Let $\mathcal{N} \subset S_X = \{x \in X\colon \|x\| = 1\}$ be ε-dense. If for all $x \in \mathcal{N}$, $B \le \|Tx\| \le A$, then*

$$\left(B - \frac{\varepsilon A}{1-\varepsilon}\right)\|x\| \le \|Tx\| \le \frac{A}{1-\varepsilon}\|x\| \quad \textit{for all } x \in X$$

This is an approximation lemma, which tells us that if we can estimate $\|Tx\|$ for just an ε-dense set of X, then we can obtain an estimate for the rest of X.

Proof. Let $x \in S_X$, then there exists $y \in \mathcal{N}$ such that $\|x-y\| \le \varepsilon$. Then

$$\|Tx\| \le \|Ty\| + \|T(x-y)\| \le A + \|T\|\,\|x-y\| \le A + \|T\|\varepsilon$$

which shows that $\|T\| \le \varepsilon\|T\| + A$ and thus $\|T\| \le \frac{A}{1-\varepsilon}$. This shows that $\|Tx\| \le A\|x\|$ for all $x \in X$. For the other direction, we have for $x \in S_X$,

$$\|Tx\| \ge \|Ty\| - \|T(x-y)\| \ge B - \|T\|\varepsilon \ge B - \frac{A\varepsilon}{1-\varepsilon}$$

and thus by scaling we have the lower bound.

□

We also have the following Lemma that gives us ε-dense sets of a particular size.

Lemma 22. *Let $\|\cdot\|$ be a norm for $\mathbb{R}^n$ and $K = \{x \in \mathbb{R}^n \colon \|x\| \le 1\}$. Then there exists $\mathcal{N} \subset K$ that is ε-dense with*

$$\#\mathcal{N} \le (3/\varepsilon)^n = e^{n\log(3/\varepsilon)}$$

Proof. Let $\mathcal{N} \le K$ be a maximal subset with respect to $\|x-y\| > \varepsilon$ for all $x, y \in \mathcal{N}$. Note that $\mathcal{N}$ is an ε-net, since otherwise if we have not covered some $z \in K$, then $\mathcal{N} \cup \{z\}$ is another subset with all points separated by ε, contradicting maximality. Furthermore, $\{x + \frac{\varepsilon}{2}K\}_{x\in\mathcal{N}}$ are disjoint sets whose union is contained in $(1+\varepsilon/2)K$. We then compare volumes:

$$\left(1+\frac{\varepsilon}{2}\right)^n \mathrm{vol}(K) = \mathrm{vol}\left(\left(1+\frac{\varepsilon}{2}\right)K\right) \ge \mathrm{vol}\left(\bigcup_{x\in\mathcal{N}} \left\{x + \frac{\varepsilon}{2}K\right\}\right) = \sum_{x\in\mathcal{N}} \mathrm{vol}\left\{x+\frac{\varepsilon}{2}K\right\} = (\#\mathcal{N})\left(\frac{\varepsilon}{2}\right)^n \mathrm{vol}(K)$$

and thus $(\#\mathcal{N}) \le \left(\frac{1+\varepsilon/2}{\varepsilon/2}\right)^n \le (3/\varepsilon)^n$.

□

From here, there are 3 steps to Dvoretsky's Theorem.

Dvoretsky Criterion

Theorem 23. (Dvoretsky Criterion) *Let $X = (\mathbb{R}^N, \|\cdot\|)$ be a N-dimensional normed space. Define*

$$\begin{aligned} L &:= \sup_{x\in\mathbb{S}^{n-1}} \|x\| \\ M &:= \int_{\mathbb{S}^{n-1}} \|x\|\, d\mu(x) \end{aligned}$$

For all $\varepsilon \in (0,1)$, there exists $C(\varepsilon) > 0$ such that for all X (any normed space!), if

$$K \le C(\varepsilon)\left(\frac{M}{L}\right)^2 N$$

then there exists a K dimensional subspace $V \subset \mathbb{R}^N$ such that for all $x \in V$,

$$(1-\varepsilon) M \|x\|_2 \leq \|x\|_X \leq (1+\varepsilon) M \|x\|_2$$

or equivalently,

$$\frac{1}{(1+\varepsilon)M}(B_2^N \cap V) \subseteq V \cap K \subseteq \frac{1}{(1-\varepsilon)M}(B_2^N \cap V)$$

(norm inequality and geometric inclusions get reversed)

Proof. Fix $V_0 \subset \mathbb{R}^N$ of dimension K. Let $\mathcal{N} \subset \mathbb{S}^{N-1} \cap V_0$ be ε-dense subset with

$$\#\mathcal{N} \leq e^{K \log(3/\varepsilon)}$$

(from Lemma 22). Let ν be the Haar measure on $O(N)$, and fix $x_0 \in \mathbb{S}^{N-1} \cap V_0$. As before, we have

$$\nu(U \in O(N) \colon |\|Ux_0\| - M| \geq \varepsilon M) = \mu\big(x \in \mathbb{S}^{N-1} \colon |\|x\| - M| \geq \varepsilon M\big)$$

since $x \mapsto \|x\|$ is a L-Lipschitz map from the assumptions, and thus the measure above is bounded by

$$\nu(U \in O(N) \colon |\|Ux_0\| - M| \geq \varepsilon M) = \mu\big(x \in \mathbb{S}^{N-1} \colon |\|x\| - M| \geq \varepsilon M\big) \leq A e^{-BN\varepsilon^2 M^2/L^2}$$

This is for a fixed x_0. Now let's vary x_0 to be elements of our ε-net $\mathcal{N}$. From the union bound, at least one of the $\#\mathcal{N}$ elements $x \in \mathcal{N}$ will fail to satisfy $|\|Ux\| - M| \geq \varepsilon M$ with probability bounded above by

$$(\#\mathcal{N}) A e^{-BN\varepsilon^2 M^2/L^2} \leq A e^{K \log(3/\varepsilon) - BN\varepsilon^2 M^2/L^2}$$

which is <1 if $K \leq c \frac{\varepsilon^2}{\log(3/\varepsilon)} \left(\frac{M}{L}\right)^2 N$, in which case there exists $U \in O(N)$ such that

$$M(1-\varepsilon) \leq \|Ux\| \leq M(1+\varepsilon) \quad \text{for all } x \in \mathcal{N}$$

Now let $V = UV_0$, and $\mathcal{N}' = U\mathcal{N}$ which is a net in $\mathbb{S}^{N-1} \cap V$. Let T be the identity mapping between $(V, \|\cdot\|_2)$ and $(V, \|\cdot\|_X)$. By approximation lemma with the operator TU, we have that for all $x \in V$,

$$\left(1 - \varepsilon - \frac{\varepsilon}{1-\varepsilon}\right) M \|x\|_2 \leq \|x\| \leq M \frac{1+\varepsilon}{1-\varepsilon} \|x\|_2$$

□

The next step is to find a normed space for which we can control M, L so that we can find high dimensional subspaces V.

Definition: An ellipsoid $\mathcal{E} \subset \mathbb{R}^n$ is an image of B_2^n under an invertible linear transformation. (Think singular value decomposition, which tells us exactly how the axes of the unit ball get rotated and stretched).

We have the following observation:

Proposition 24. *If $X = (\mathbb{R}^n, \|\cdot\|)$, and $K = \{x \in \mathbb{R}^n \colon \|x\| \leq 1\}$, then there exists an ellipsoid with maximum volume contained in K.*

Actually, the ellipsoid is unique, and is called the John ellipsoid. We will not prove this now, however.

Proof. This follows by a compactness argument. We want $T: B_2^n \to \mathbb{R}^n$ with $TB_2^n \subset K$, which means that $\|T\|_{l_2^n \to X} \leq 1$ (the operator norm). The set of such T form a compact set, since it is the unit ball of some n^2-dimensional normed space (in finite dimensions, the unit ball is compact). Therefore since the volume $|\det(T)|$ is a continuous function, it achieves its maximum on this compact set. This shows the existence of an ellipsoid of maximum volume contained in K. □

Next time we will prove the final ingredient.

Week 4 (9/28/2010)

Remark 25. Note that for $\mathcal{E} \subset \mathbb{R}^n$, $\mathcal{E} = TB_2^n$ is the unit ball of the norm $\|X\|_{\mathcal{E}} = \|T^{-1}x\|_2$, and in fact the norm comes from a Hilbert space with an inner product defined by

$$\langle x, y\rangle_{\mathcal{E}} := \langle T^{-1}x, T^{-1}y\rangle$$

Dvoretsky-Rogers Lemma

Lemma 26. (Dvoretsky-Rogers) *Let $X = (\mathbb{R}^n, \|\cdot\|)$ be a normed space, $K = \{x: \|x\| \leq 1\}$, and $\mathcal{E}$ an ellipsoid of maximal volume contained in K. Then there exists an orthonormal basis $\{x_1, ..., x_n\}$ of $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_{\mathcal{E}})$ such that*

$$\|x_i\| \geq \frac{1}{2}\sqrt{1 - \frac{i-1}{n}}$$

Afterwards we will analyze what M, L are for this space, and the result will soon follow.

Proof. Construct $x_1, ..., x_n$ inductively. Take x_1 to be an arbitrary point such that $\|x\|_{\mathcal{E}} = \|x\| = 1$. Since $\mathcal{E}$ is the maximal ellipsoid contained in K, there must exist such a point. Now suppose we have defined $x_1, ..., x_{i-1}$. Let x_i be a point of $\mathcal{E}$ such that

$$\|x_i\| = \max\{\|x\|: \langle x, x_j\rangle_{\mathcal{E}} = 0, j = 1, ..., i-1, x \in \mathcal{E}\}$$

By construction, this is already an orthonormal basis with respect to $\mathcal{E}$. Now we try to find a bound for $\|x_i\|$. Note that for $x \in \operatorname{span}\{x_i, ..., x_n\}$, we have that $\langle x, x_j\rangle_{\mathcal{E}} = 0$ for $j = 1, ..., i-1$ and thus

$$\left\|\frac{x}{\|x\|_{\mathcal{E}}}\right\| \leq \|x_i\|$$

by definition of $\|x_i\|$. Now fix i, and let $a, b > 0$, to be chosen later. Define a new ellipsoid $\mathcal{E}'$:

$$\mathcal{E}' := \left\{\sum_{j=1}^{i-1} a a_j x_j + \sum_{j=i}^{n} b a_j x_j: \sum_{j=1}^{n} a_j^2 \leq 1\right\}$$

That is, $\mathcal{E}' = T\mathcal{E}$ with $T = \begin{pmatrix} a & & & & & \\ & \ddots & & & & \\ & & a & & & \\ & & & b & & \\ & & & & \ddots & \\ & & & & & b \end{pmatrix}$ where a appears $i-1$ times and b appears $n-i+1$ times.

Note that if $\sum_{j=1}^n b_j x_j \in \mathcal{E}'$ then $\sum_{j=1}^{i-1}\left(\frac{b_j}{a}\right)^2 + \sum_{j=i}^n \left(\frac{b_j}{b}\right)^2 \leq 1$. Now:

$$\begin{aligned}
\left\|\sum_{j=1}^n b_j x_j\right\| &\leq \left\|\sum_{j=1}^{i-1} b_j x_j\right\| + \left\|\sum_{j=i}^n b_j x_j\right\| \\
&\leq \left\|\sum_{j=1}^{i-1} b_j x_j\right\|_{\mathcal{E}} + \|x_i\| \left\|\sum_{j=i}^n b_j x_j\right\|_{\mathcal{E}} \\
&= \left(\sum_{j=1}^{i-1} b_j^2\right)^{1/2} + \|x_i\| \left(\sum_{j=i}^n b_j^2\right)^{1/2} \\
&= a\left(\sum_{j=1}^{i-1}\left(\frac{b_j}{a}\right)^2\right)^{1/2} + b\,\|x_i\|\left(\sum_{j=i}^n\left(\frac{b_j}{b}\right)^2\right)^{1/2} \\
&\leq \sqrt{a^2 + b^2\|x_i\|^2}\left(\sum_{j=1}^{i-1}\left(\frac{b_j}{a}\right)^2 + \sum_{j=i}^n\left(\frac{b_j}{b}\right)^2\right) \\
&\leq \sqrt{a^2 + b^2\|x_i\|^2}
\end{aligned}$$

where we used the fact that $\sum_{j=i}^n b_j x_j \in \{x_1, ..., x_{i-1}\}^{\perp \varepsilon}$ and the fact that $\mathcal{E} \subset K$ so $\|x\| \leq \|x\|_{\mathcal{E}}$ in the second inequality. The equalities in the middle follow from $\{x_1, ..., x_n\}$ bein an orthonormal basis with respect to $\mathcal{E}$ (Parseval). The second to last inequality follows from $a\lambda + b\mu \leq (a^2 + b^2)^{1/2}(\lambda^2 + \mu^2)^{1/2}$ (Cauchy-Schwarz).

So $\mathcal{E}' \subset K$ if $a^2 + b^2\,\|x_i\|^2 \leq 1$. In this case, by the maximality of $\mathcal{E}$, we have that

$$a^{i-1} b^{n-i+1} \operatorname{vol}(\mathcal{E}) = \operatorname{vol}(\mathcal{E}') \leq \operatorname{vol}(\mathcal{E})$$

and this implies that $a^{i-1} b^{n-i+1} \leq 1$ as long as $a^2 + b^2\,\|x_i\|^2 \leq 1$. Now let's choose a, b to maximize

$$\max\left\{a^{i-1} b^{n-i+1} \colon a^2 + b^2\,\|x_i\|^2 \leq 1\right\}$$

Lagrange Multipliers: Equivalent to maximize $(i-1)\log a + (n-i+1)\log b$, so we get the equations $\frac{i-1}{a} = \lambda 2a$, $\frac{n-i+1}{b} = \lambda 2b\,\|x_i\|^2$.

$$\lambda = \frac{i-1}{2a^2} = \frac{n-i+1}{2b^2\,\|x_i\|^2}$$

or $b^2 = \frac{a^2\,(n-i+1)}{(i-1)\|x_i\|^2}$, so plugging to constraint gives

$$a^2\left(\frac{n}{i-1}\right) = 1$$

i.e. $a = \sqrt{\frac{i-1}{n}}$ and $b = \frac{\sqrt{1-a^2}}{\|x_i\|} = \sqrt{\frac{n-i+1}{n\,\|x_i\|^2}}$ (from constraint).

Using this choice of a, b, we have that

$$\left(\frac{i-1}{n}\right)^{\frac{i-1}{2}}\left(\frac{n-i+1}{n\,\|x_i\|^2}\right)^{\frac{n-i+1}{2}} \leq 1$$

and therefore

$$\|x_i\| \geq \left(1 + \frac{n-i+1}{i-1}\right)^{-\frac{i-1}{2(n-i+1)}} \sqrt{1 - \frac{i-1}{n}} \geq \frac{1}{2}\sqrt{1 - \frac{i-1}{n}}$$

noting that we can bound $\left(1+\frac{1}{x}\right)^x \leq 4$. Note that since the limit as $x \to \infty$ tends to e and the limit as $x \to 0$ tends to 1, we can already bound by some constant. Can be more precise as well:

For $x < 1$, we have the estimate

$$x \log\left(1+\frac{1}{x}\right) \leq \frac{\log\left(\frac{2}{x}\right)}{\frac{1}{x}}$$

Now $\frac{\log 2u}{u}$ achieves its maximum when $\frac{\frac{1}{2} - \log 2u}{u^2} = 0$ or at $u = \frac{1}{2}e^{1/2}$ with the value $e^{-1/2}$. Thus

$$x \log\left(1+\frac{1}{x}\right) \leq \frac{\log \frac{2}{x}}{\frac{1}{x}} \leq e^{-1/2} \leq 1$$

Thus $\left(1+\frac{1}{x}\right)^x \leq e$ for $x < 1$.

For other x, we have a bound using power series:

$$x \log\left(1+\frac{1}{x}\right) = x\left(\frac{1}{x} - \frac{2}{x^2} + \frac{3}{x^3} - \ldots\right) \leq x\left(\frac{1}{x}\right) = 1$$

so long as $\frac{k}{x^k} \geq \frac{k+1}{x^{k+1}}$ for all $k \geq 1$, or $x \geq \frac{k+1}{k} = 1 + \frac{1}{k} \geq 2$ (alternating series estimate), and thus $\left(1+\frac{1}{x}\right)^x \leq e$ for $x \geq 2$.

Now for $1 \leq x < 2$, we have

$$\left(1+\frac{1}{x}\right)^x \leq \left(1+\frac{1}{x}\right)^2 \leq 4$$

Thus $\left(1+\frac{1}{x}\right)^x \leq 4$ and $\left(1+\frac{1}{x}\right)^{-x/2} \geq \frac{1}{2}$.

This proves the result. □

Theorem 27. (Ellipsoid Version of Dvoretzky's Theorem) *For all $\varepsilon > 0$, there exists $C(\varepsilon) > 0$ such that the following holds:*

If $\|\cdot\|$ is a norm on $\mathbb{R}^n$ and K is the corresponding unit ball, then there exists an ellipsoid $\mathcal{E}_0 \subset \mathbb{R}^n$ and a linear subspace $V \subset \mathbb{R}^n$ such that

1. *There exists $r > 0$ for which*

$$r(\mathcal{E}_0 \cap V) \subseteq K \cap V \subseteq (1+\varepsilon)\, r(\mathcal{E}_0 \cap V)$$

2. $\dim V \geq C(\varepsilon) \log n$

Proof. Let $\mathcal{E}$ be an ellipsoid of maximum volume, with $\mathcal{E} = TB_2^n$. First we note that it suffices to prove the result for the case when $T = \mathrm{Id}$ or $\mathcal{E} = B_2^n$:

This is because B_2^n is the ellipsoid of maximal volume in $T^{-1}K$. Now suppose we proved the result for B_2^n. This means that there exists $V \subset \mathbb{R}^n$ with $\dim V \geq C(\varepsilon) \log n$ and

$$r(B_2^n \cap V) \subseteq K \cap V \subseteq (1+\varepsilon)\, r(B_2^n \cap V)$$

This means that

$$r(\mathcal{E} \cap (TV)) \subseteq K \cap (TV) \subseteq (1+\varepsilon) r(\mathcal{E} \cap (TV))$$

and TV is the desired subspace (with the same dimension since T is invertible).

Thus, without loss of generality we will assume that B_2^n is the ellipsoid of maximum volume of K. By Dvoretsky's Criterion (Theorem 23), there exists a subspace $V\subset\mathbb{R}^n$ with $\dim V\geq C(\varepsilon)M^2n$ where

$$\frac{1}{(1+\varepsilon)M}(V\cap B_2^n)\subseteq V\cap K\subseteq\frac{1}{(1-\varepsilon)M}(V\cap B_2^n)$$

Note that in our case $L=\max_{x\in\mathbb{S}^{n-1}}\|x\|=1$. We will show that $M\gtrsim\sqrt{\frac{\log n}{n}}$, which proves the result.

Claim: $M\geq c\sqrt{\frac{\log n}{n}}$.

By Dvoretsky-Rogers (Lemma 26) there exists an orthonormal basis $x_1,\ldots,x_n$ with

$$\|x_i\|\geq C \text{ for } i\leq\frac{n}{2}$$

(this is a weaker statement than the Lemma). Computing:

$$\begin{aligned}
M &= \int_{\mathbb{S}^{n-1}}\|x\|\,d\mu(x)\\
\text{(rotation)} &= \int_{\mathbb{S}^{n-1}}\left\|\sum a_i x_i\right\|d\mu(a)\\
\text{(same distribution)} &= \int_{\mathbb{S}^{n-1}}\left\|\sum \varepsilon_i a_i x_i\right\|d\mu(a)\quad\text{for each fixed } \varepsilon_i\in\{\pm1\}\\
\left(M=\frac{1}{2^n}\sum_{\varepsilon}M\right) &= \int_{\mathbb{S}^{n-1}}\frac{1}{2^n}\sum_{\varepsilon\in\{\pm1\}^n}\left\|\sum_{i=1}^n\varepsilon_i a_i x_i\right\|d\mu(a)
\end{aligned}$$

Here we note that

$$\frac{1}{2^n}\sum_{\varepsilon\in\{\pm1\}^n}\left\|\sum_{i=1}^n\varepsilon_i a_i x_i\right\|\geq\max\|a_i x_i\|$$

because

$$\begin{aligned}
\frac{1}{2^n}\sum_{\varepsilon\in\{\pm1\}^n}\left\|\sum_{i=1}^n\varepsilon_i a_i x_i\right\| &= \frac{1}{2^{n-1}}\sum_{\substack{\varepsilon_1,\ldots,\varepsilon_{j-1},\\ \varepsilon_{j+1},\ldots,\varepsilon_n=\pm1}}\frac{\left\|a_jx_j+\sum_{i\neq j}\varepsilon_i a_i x_i\right\|+\left\|-a_jx_j+\sum_{i\neq j}\varepsilon_i a_i x_i\right\|}{2}\\
&\geq \frac{1}{2^{n-1}}\sum\|a_j x_j\|\\
&= \|a_j x_j\|
\end{aligned}$$

where we used the triangle inequality $\frac{\|x+y\|+\|x-y\|}{2}\geq\left\|\frac{x+y}{2}-\frac{x-y}{2}\right\|=\|y\|$.

This means that continuing the computation above,

$$\begin{aligned}
M &= \int_{\mathbb{S}^{n-1}}\frac{1}{2^n}\sum_{\varepsilon\in\{\pm1\}^n}\left\|\sum_{i=1}^n\varepsilon_i a_i x_i\right\|d\mu(a)\\
&\geq \int_{\mathbb{S}^{n-1}}\max_{1\leq i\leq n}(|a_i|\,\|x_i\|)\,d\mu(a)\\
&\geq \int_{\mathbb{S}^{n-1}}\max_{1\leq i\leq n/2}(|a_i|\,\|x_i\|)\,d\mu(a)\\
&\geq C\int_{\mathbb{S}^{n-1}}\max_{1\leq i\leq n/2}|a_i|\,d\mu(a)
\end{aligned}$$

We want to show that

$$\int_{\mathbb{S}^{n-1}} \max_{1\le i\le n/2} |a_i|\, d\mu(a) \ge \sqrt{\frac{\log n}{n}}$$

and we will be finished. This will follow from two facts:

1. Let γ_n be the standard Gaussian on $\mathbb{R}^n$, with $d\gamma_n(x) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|_2^2/2}$. If $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is integrable and homogeneous of order 1, i.e. $f(rx) = r f(x)$ for all $r > 0$, then

$$\int_{\mathbb{S}^{n-1}} f(x)\, d\mu(x) = \frac{\int_{\mathbb{R}^n} f(x)\, d\gamma_n}{\int_{\mathbb{R}^n} \|x\|_2\, d\gamma_n} \approx \frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} f(x)\, d\gamma_n(x)$$

2. Gaussian tail bound:

$$\gamma_1(\{|x| > t\}) = \frac{2}{\sqrt{2\pi}} \int_t^\infty e^{-u^2/2} du \le \frac{1}{10} e^{-t^2/2}$$

First assuming these facts: we have that

$$\int_{\mathbb{S}^{n-1}} \max_{1\le i\le n/2} |a_i|\, d\mu(a) \approx \frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \max_{1\le i\le n/2} |x_i|\, d\gamma_n(x)$$

We then need to show that $\int_{\mathbb{R}^n} \max_{1\le i\le n/2} |x_i|\, d\gamma_n \gtrsim \sqrt{\log n}$. Let $t > 1$. Then

$$\gamma_n\left(\max_{1\le i\le n/2} |x_i| \ge t \right) = 1 - (\gamma_1(|x_1| < t))^{n/2} \ge 1 - \left(1 - \frac{1}{10} e^{-t^2/2} \right)^{n/2}$$

using the Gaussian tail bound here. Choosing $t = \sqrt{2\log(n)}$, we then have that this is bounded below by

$$\ge 1 - \left(1 - \frac{1}{10n} \right)^{n/2} \ge C$$

Thus,

$$\int_{\mathbb{R}^n} \max_{1\le i\le n/2} |x_i|\, d\gamma_n \ge \int_{\max_{1\le i\le n/2} |x_i| \ge \sqrt{2\log n}} |x_i|\, d\gamma_n \ge C\sqrt{2\log n}$$

as desired.

$\square$

Proof. (of Gaussian Integral Facts)

(1) This is a direct computation in polar coordinates.

$$\begin{aligned} \int_{\mathbb{R}^n} f(x)\, d\gamma_n &= c_n \int_0^\infty r^{n-1} e^{-r^2/2} \int_{\mathbb{S}^{n-1}} f(rx)\, d\mu(x)\, dr \\ &= c_n \left(\int_0^\infty r^n e^{-r^2/2}\, dr \right) \left(\int_{\mathbb{S}^{n-1}} f(x)\, d\mu(x) \right) \end{aligned}$$

We note that if we look at $f(x) = \|x\|_2$, the above computation shows that

$$c_n \int_0^\infty r^n e^{-r^2/2}\, dr = \int_{\mathbb{R}^n} \|x\|_2\, d\gamma_n$$

We claim that $\int_{\mathbb{R}^n} \|x\|_2 \, d\gamma_n \approx \sqrt{n}$. With the second moment, we have

$$\int_{\mathbb{R}^n} \|x\|_2^2 \, d\gamma_n(x) = \int_{\mathbb{R}^n} \sum_{i=1}^n x_i^2 \, d\gamma_n(x) = \sum_{i=1}^n \int_{\mathbb{R}} x_i^2 \, d\gamma_1(x_i) = n$$

Also,

$$\begin{aligned}
\int_{\mathbb{R}^n} \|x\|_2^4 \, d\gamma_n(x) &= \int_{\mathbb{R}^n} \left(\sum_{i,j} x_i^2 x_j^2 \right) d\gamma_n(x) \\
&= \sum_{i \neq j} \int_{\mathbb{R}} x_i^2 \, d\gamma_1(x_i) \int_{\mathbb{R}} x_j^2 \, d\gamma_1(x_j) + \sum_i \int_{\mathbb{R}} x_i^4 \, d\gamma_1(x_i) \\
&= (n^2 - n) + Cn \\
&\lesssim n^2
\end{aligned}$$

If we let $Z = \|x\|_2$, we have shown that $\left(\mathbb{E} Z^4\right)^{1/4} \lesssim \sqrt{\mathbb{E} Z^2}$. Then using Jensen and Holder as we did in the proof of Theorem 19, we have

$$\sqrt{\mathbb{E} Z^2} \geq \mathbb{E} Z \gtrsim \sqrt{\mathbb{E} Z^2}$$

and thus $\mathbb{E} Z = \int_{\mathbb{R}^n} \|x\|_2 \, d\gamma_n \sim \sqrt{n}$.

(2) To show the Gaussian tail estimate,

$$\begin{aligned}
\int_t^\infty e^{-u^2/2} \, du &= \int_t^\infty \frac{1}{u} \cdot u e^{-u^2/2} \, du \\
&= -\frac{1}{u} e^{-u^2/2} \Big|_t^\infty - \int_t^\infty \frac{1}{u^2} e^{-u^2/2} \, du
\end{aligned}$$

This means that

$$\int_t^\infty \left(1 + \frac{1}{u^2}\right) e^{-u^2/2} \, du = \frac{1}{t} e^{-t^2/2}$$

$$\int_t^\infty e^{-u^2/2} \, du \leq \int_t^\infty \left(1 + \frac{1}{u^2}\right) e^{-u^2/2} \, du = \frac{1}{t} e^{-t^2/2}$$

and thus

$$\int_t^\infty e^{-u^2/2} \, du \leq \frac{1}{10} e^{-t^2/2}$$

for sufficiently large t (in fact, $t > 10$).

□

Recall that in Dvoretsky's Theorem (Theorem 7) we wanted to use B_2^n rather than the ellipsoid $\mathcal{E}$. The final step is to show that there is a further subspace for which Dvoretsky will hold with $\mathcal{E}$ replaced by B_2^n.

Lemma 28. *Let $\mathcal{E} \subset \mathbb{R}^n$ be an ellipsoid. Then there exists a subspace $V \subset \mathbb{R}^n$ and a number $r > 0$ such that*

1. $\dim V \geq \frac{n}{16}$
2. $\mathcal{E} \cap V = r(B_2^n \cap V)$

Proof. Assume that n is divisible by 4, and find V with $\dim V \geq \frac{n}{4}$. We will start by constructing $x_1, \ldots, x_{n/2}$ in $\mathbb{R}^n$ such that

1. $\|x_i\|_2 = 1$ for all i
2. For all $i \neq j$, $\langle x_i, x_j \rangle = \langle x_i, x_j \rangle_{\mathcal{E}} = 0$

Start with an arbitrary $\|x_1\|_2 = 1$. Assume we have defined $x_1 \ldots, x_i$. Look at

$$\mathcal{U} = \left\{ x \in \mathbb{R}^n, \langle x, x_i \rangle = \langle x, x_i \rangle_{\mathcal{E}} = 0, \text{ for all } j \leq i \right\}$$

Note that $\dim \mathcal{U} \geq n - 2i$. (There are $2i$ linear constraints imposed on $\mathcal{U}$) So if $i < n/2$, then $\mathcal{U} \neq \{0\}$ and we can find $x_{i+1} \neq 0$ in $\mathcal{U}$, and normalize it.

Now assume that $\|x_1\|_{\mathcal{E}} \geq \ldots \geq \|x_{n/2}\|_{\mathcal{E}}$. Choose λ such that

$$\|x_{n/4}\|_{\mathcal{E}} \geq \lambda \geq \|x_{n/4+1}\|_{\mathcal{E}}$$

Then there exists a $\lambda_1 \in (0, 1)$ such that

$$y_1 := \lambda_1 x_1 + \sqrt{1 - \lambda_1^2}\, x_{n/2}$$

satisfies $\|y_1\|_{\mathcal{E}} = \lambda$. This follows simply from the fact that

$$\|y_1\|_{\mathcal{E}} = \sqrt{\lambda_1^2 \|x_1\|_{\mathcal{E}}^2 + (1 - \lambda_1^2) \|x_{n/2}\|_{\mathcal{E}}^2}$$

Likewise, we can find $\lambda_2 \in (0, 1)$ such that

$$y_2 := \lambda_2 x_2 + \sqrt{1 - \lambda_2^2}\, x_{n/2-1}$$

satisfies $\|y_2\|_{\mathcal{E}} = \lambda$. Continue this to obtain $y_1, \ldots, y_{n/4}$. We now have a collection for which $\|y_i\|_2 = 1$, $\|y_i\|_{\mathcal{E}} = \lambda$ and $\langle y_i, y_j \rangle = \langle y_i, y_j \rangle_{\mathcal{E}} = \delta_{ij}$. Take $V = \operatorname{span}\{y_1 \ldots, y_{n/4}\}$. Then $x = \sum_{i=1}^{n/4} a_i v_i$ and

$$\|x\|_2 = \sqrt{\sum_{i \leq n/4} a_i^2}, \quad \|x\|_{\mathcal{E}} = \lambda \sqrt{\sum_{i \leq n/4} a_i^2}$$

which is what we wanted.

□

Remark 29. Note above that we may even take V to be a subspace of any other linear subspace W, since in the construction we have started with an arbitrary subspace of a specified dimension.

This finishes the proof of Dvoretsky's Theorem. We use the Ellipsoid version to find $\mathcal{E}$ and V, and then find a further subspace $W \subset V$ so that $\mathcal{E} \cap W$ becomes a slice of B_2^n.

Sharpness of Dvoretsky

Let us look at what happens when we look at the l^∞ norm.

Lemma 30. *Let $\varepsilon \in (0, 1)$. If $V \subset \mathbb{R}^n$ is a subspace such that for all $x \in V$,*

$$r \|x\|_\infty \leq \|x\|_2 \leq (1 + \varepsilon) r \|x\|_\infty$$

Then

$$\dim V \leq \frac{C}{\log(1/\varepsilon)} \log n$$

This shows that the log n factor is sharp. In our proof, we obtained the ε-dependent constant $C(\varepsilon) = \frac{\varepsilon^2}{\log(1/\varepsilon)}$ (Proof of Theorem 23). It is a known result (Schechtman) that we can obtain

$$C(\varepsilon) \sim \frac{c\varepsilon}{(\log 1/\varepsilon)^2}$$

There is a gap of a factor of $\varepsilon/(\log 1/\varepsilon)$, and closing the gap is currently an open problem.

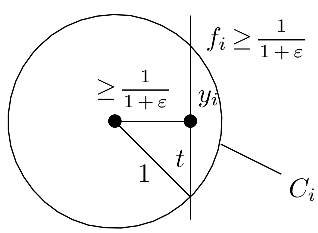
Proof. (of Lemma) Write $\dim V = k$. Then if $x \in V$, we have that

$$\max_{1 \leq i \leq n} |f_i(x)| \leq \|x\|_2 \leq \max_{1 \leq i \leq n} |f_i(x)| (1 + \varepsilon)$$

where $f_i(x) = \langle x, r e_i \rangle$, n linear functionals on $\mathbb{R}^n$. We will show that n must be big. The left inequality says that $\|f_i\| \leq 1$ for all i, and the right inequality says that for all $x \in S^{k-1} = \mathbb{S}^{n-1} \cap V$, there exists i such that $|f_i(x)| \geq \frac{1}{1+\varepsilon}$. Then if we consider the cap

$$C_i := \left\{ x \in S^{k-1} \colon f_i(x) \geq \frac{1}{1+\varepsilon} \right\}$$

then $\pm C_i$ covers all of S^{k-1}. We can cover each cap C_i by a ball using Pythagorean theorem:



(Note $\|y_i\|_2 \geq |f_i(y_i)| \geq \frac{1}{1+\varepsilon}$) Pythagorean theorem gives that

$$t \leq \sqrt{1 - \left(\frac{1}{1+\varepsilon}\right)^2} \leq 2\sqrt{\varepsilon}$$

and that $C_i \subset B(y_i, 2\sqrt{\varepsilon})$ (note in the direction of y_i that $1 - \frac{1}{1+\varepsilon} = \frac{\varepsilon}{1+\varepsilon} \leq \varepsilon < 2\sqrt{\varepsilon}$). This implies that we have $y_1, \ldots, y_{2n}$ of length $\geq \frac{1}{1+\varepsilon}$, yet

$$S^{k-1} \subseteq \bigcup_{i=1}^{2n} B(y_i, 2\sqrt{\varepsilon})$$

Now we use a volume argument. First we cover a tubular region by enlarging the balls:

$$(1 - \sqrt{\varepsilon}, 1 + \sqrt{\varepsilon}) S^{k-1} \subseteq \bigcup_{i=1}^{2n} B(y_i, 4\sqrt{\varepsilon})$$

Thus

$$\left[(1+\sqrt{\varepsilon})^k-(1-\sqrt{\varepsilon})^k\right]\mathrm{vol}(B_2^k)\leq 2n(4\sqrt{\varepsilon})^k\,\mathrm{vol}(B_2^k)$$

Since $(1+\sqrt{\varepsilon})^k-(1-\sqrt{\varepsilon})^k\geq\frac{1}{2}k\sqrt{\varepsilon}$, we have that

$$\frac{k}{4^k}\lesssim n\sqrt{\varepsilon}^{k-1}$$

$$\log k-k\log 4\lesssim\log n+k\log\sqrt{\varepsilon}-\log\sqrt{\varepsilon}$$

$$\log k-k\log 4+\frac{1}{2}\log 1/\varepsilon\lesssim\log n+\frac{1}{2}\log 1/\varepsilon$$

$$k\lesssim\frac{\log n}{\log 1/\varepsilon}$$

□

Week 5 (10/5/2010)

Notation: For $D\geq 1$, if X,Y are normed spaces, we will write $X\overset{D}{\hookrightarrow}Y$ if there exists a one-to-one linear map $T\colon X\to Y$ such that $\|T\|\,\|T^{-1}\|\leq D$.

Summary of Dvoretsky's Theorem

Here we state the various ways of writing the conclusion of Dvoretsky's Theorem (Theorem 7):

For every $\varepsilon\in(0,1)$, there exists a constant $C(\varepsilon)$ such that the following statement holds:

If $K\subset\mathbb{R}^n$ is a centrally symmetric convex body (compact, with nonempty interior), then there exists a linear subspace $V\subset\mathbb{R}^n$ such that

1. Exists $r>0$ such that

$$r(B_2^n\cap V)\subset(K\cap V)\subset(1+\varepsilon)r(B_2^n\cap V)$$

2. $\dim V\geq c(\varepsilon)\log n$

$B_2^n=\{x\in\mathbb{R}^n\colon\sum x_i^2\leq 1\}$ is the Euclidean unit ball.

If $\|\cdot\|$ is the norm associated to K, then we can write the first statement as

$$\frac{1}{(1+\varepsilon)r}\|x\|_2\leq\|x\|\leq\frac{1}{r}\|x\|_2\ \text{ for all } x\in V$$

If we let T be the mapping from $l_2^k=(V,\|\cdot\|_2)\to(\mathbb{R}^n,\|\cdot\|)=X$, where $k\geq c(\varepsilon)\log n$, the right inequality says that $\|T\|\leq\frac{1}{r}$ whereas the left inequality says that $\|T^{-1}\|\leq(1+\varepsilon)r$ so that $\|T\|\,\|T^{-1}\|\leq 1+\varepsilon$, and with our notation we have that

$$l_2^k\overset{1+\varepsilon}{\hookrightarrow}X$$

Last time we showed in Lemma 30 that if we have $l_2^k \overset{1+\varepsilon}{\hookrightarrow} l_\infty^n$, then $k \le \frac{C}{\log(1/\varepsilon)} \log n$.

Exercise 3. Prove that there exists $k \gtrsim \frac{1}{\log(1/\varepsilon)} \log n$ such that $l_2^k \overset{1+\varepsilon}{\hookrightarrow} l_\infty^n$. In other words, the result of the lemma is sharp. (Hint: Reverse the argument we used in the Lemma)

The goal is to find $U \in O(n)$ such that if we choose $y_i = U e_i$, then the union of $B(\pm y_i, 2\sqrt{\varepsilon})$ covers $\mathbb{S}^{n-1}$.

Now we already have that $\|x\|_\infty \le \|x\|_2$, and to prove that $\|x\|_\infty \ge \frac{1}{1+\varepsilon}\|x\|_2$, we can use this cover. For $\|x\| = 1$, i.e. $x \in \mathbb{S}^{n-1}$, we have that there is some y_i for which $\|x - y_i\|_2 \le 2\sqrt{\varepsilon}$. This means that

$$\begin{aligned}
4\varepsilon &\ge \|x - y_i\|_2^2 \\
4\varepsilon &= 1 - \frac{2}{1+\varepsilon}\langle x, e_i\rangle + \frac{1}{(1+\varepsilon)^2} \\
x_i &\ge \frac{1}{2}\left[1 + \varepsilon + \frac{1}{1+\varepsilon} - 4\varepsilon(1+\varepsilon)\right] \sim \frac{1}{1+\varepsilon'}
\end{aligned}$$

and thus $\|x\|_\infty \ge \frac{1}{1+\varepsilon'}$, which proves the result.

However, not sure how to show the existence of such a U by reversing the argument.

This tells us how small k can be so that an embedding exists.

Let us now look at how *large* k can be for specific spaces $X = l_p^n$ (note already that it is trivial to embed $l_2^k \overset{1}{\hookrightarrow} l_2^n$ for k all the way up to n with any orthognal projection)

We return to the Dvoretsky Criterion (Theorem 23), which states that if $\|\cdot\|$ is a norm on $\mathbb{R}^n$ and K is the corresponding unit ball, where $\|x\| \le L\,\|x\|_2$ for all $x \in \mathbb{R}^n$, then there exists $k \ge C(\varepsilon)(M/L)^2 n$ such that

$$l_2^k \overset{1+\varepsilon}{\hookrightarrow} (\mathbb{R}^n, \|\cdot\|)$$

where $M = \int_{\mathbb{S}^{n-1}} \|x\|\, d\mu(x)$.

It turns out that this is good enough to obtain sharp results for the l_p^n spaces, for $1 \le p < \infty$. Let us compute the different values of L, M.

- L .

  For $p \ge 2$, we note that $\|x\|_p \le \|x\|_2$, which implies that $L \le 1$. By homogeneity it suffices to show the result for when $\|x\|_2 \le 1$. In this case, we must have $|x_i| \le 1$ as well, and thus

  $$\sum |x_i|^p \le \sum |x_i|^2 = 1$$

  so that $\|x\|_p \le 1$.

  For $p \le 2$, we have Hölder with $2/p, \frac{1}{1-p/2}$:

  $$\|x\|_p^p = \sum_{i=1}^n |x_i|^{2\cdot\frac{p}{2}} \cdot 1 \le \left(\sum_{i=1}^n |x_i|^2\right)^{p/2} \cdot n^{1-p/2} = \|x\|_2^p\, n^{1-p/2}$$

  Thus $\|x\|_p \le n^{\frac{1}{p}-\frac{1}{2}}\|x\|_2$ so that $L \le n^{\frac{1}{p}-\frac{1}{2}}$

  In summary,

  $$L \le \begin{cases} 1 & p \ge 2 \\ n^{\frac{1}{p}-\frac{1}{2}} & p \le 2 \end{cases}$$

- M. We will just do the polar coordinate computation as in the proof of Theorem 27 (Ellipsoid version of Dvoretsky) or in Exercise 2.

$$\begin{aligned} M = \int_{\mathbb{S}^{n-1}} \|x\|_p \, d\mu(x) &\sim \frac{1}{\sqrt{n}} \int_{\mathbb{R}^n} \|x\|_p \, d\gamma_n(x) \\ \text{Minkowski} &\geq \frac{1}{\sqrt{n}} \left\| \int_{\mathbb{R}^n} |x| \, d\gamma_n(x) \right\|_p \\ &= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n \left| \int_{\mathbb{R}^n} |x_i| \, d\gamma_n(x) \right|^p \right)^{1/p} \\ &\gtrsim n^{\frac{1}{p} - \frac{1}{2}} \end{aligned}$$

Above we used that $\int_{\mathbb{R}^n} |x_i| \, d\gamma_n(x) = C$ for all i, by rotational invariance of the Gaussian measure.

This means that applying the Dvoretsky criteria gives that for all ε, $\left(k \gtrsim C(\varepsilon) \left(\frac{M}{L} \right)^2 n \right)$

$$l_2^k \xhookrightarrow{1+\varepsilon} l_p^n \text{ for some } k \gtrsim \begin{cases} n^{2/p} & p > 2 \\ n & 1 \leq p \leq 2 \end{cases}$$

In particular for $p \leq 2$ we know we can achieve embeddings for l_2^k into l_p^n for k all the way up to some proportion of n. This proves Theorem 1 that we presented in the overview.

Exercise 4. Show that

$$\int_{\mathbb{R}^n} \|x\|_p \, d\gamma_n(x) \gtrsim \sqrt{p} \, n^{1/p} \text{ for } p \geq 2$$

Thus, for $p \geq 2$, get $k \gtrsim p n^{2/p}$.

$$\begin{aligned} \int_{\mathbb{R}^n} \|x\|_p^p \, d\gamma_n(x) &= \int_{\mathbb{R}^n} \left(\sum_{i=1}^n |x_i|^p \right) d\gamma_n(x) \\ &= \sum_{i=1}^n \int_{\mathbb{R}_n} |x_i|^p \, d\gamma_n(x) \\ &= \frac{2n}{\sqrt{2\pi}} \int_0^\infty x^p e^{-x^2/2} \, dx \\ &= \frac{2n}{\sqrt{2\pi}} \int_0^\infty 2^{\frac{p-1}{2}} u^{\frac{p+1}{2} - 1} e^{-u} \, du \\ &= \frac{n 2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2} \right) \\ &\sim n (cp)^{p/2} \end{aligned}$$

This computation just shows

$$\left(\int_{\mathbb{R}^n} \|x\|_p \, d\gamma_n \right)^{p \cdot 1/p} \leq \left(\int_{\mathbb{R}^n} \|x\|_p^p \right)^{1/p} \sim n^{1/p} \sqrt{p}$$

for any p. The reverse can be proved by mimicking the idea in the proof of the Johnson Lindenstrauss Lemma (Theorem 19). We first compute the $2p$-th power:

$$\begin{aligned} \int_{\mathbb{R}^n} \|x\|_p^{2p} \, d\gamma_n(x) &= \int_{\mathbb{R}^n} \left(\sum_{i=1}^n |x_i|^p \right)^2 d\gamma_n(x) \\ &= \sum_{i,j} \int_{\mathbb{R}^n} |x_i|^p \, |x_j|^p \, d\gamma_n(x) \\ &= \sum_{i=1}^n \int_{\mathbb{R}^n} |x_i|^{2p} \, d\gamma_n(x) + \sum_{i \neq j} \int_{\mathbb{R}^n} |x_i|^p |x_j|^p \, d\gamma_n(x) \\ &= n (c 2p)^p + \sum_{i \neq j} \left(\int_{\mathbb{R}} |x| \, d\gamma_1(x) \right)^2 \\ &\sim n^2 (cp)^p \end{aligned}$$

Comparing this with the p-th power, we have that letting $Z = \|x\|_p$,

$$\mathbb{E}[Z^p] \sim \mathbb{E}[Z^{2p}]^{1/2}$$

We will use this to show that $\mathbb{E}[Z^p] \sim \mathbb{E}[Z]^p$. We already showed with convexity that $\mathbb{E}[Z^p] \gtrsim \mathbb{E}[Z]^p$. To show the other direction, we find an $a < p$ for which

$$\mathbb{E}[Z^p] = \mathbb{E}[Z^a Z^{p-a}] \leq \mathbb{E}[Z^{2p}]^{a/2p} \mathbb{E}[Z]^{p-a}$$

where $\frac{a}{2p} + p - a = 1$ (applying Hölder). This means $a = \frac{p-1}{1 - \frac{1}{2p}} = p\frac{2p-2}{2p-1}$, which is always less than p. Now with this choice of a, using the relationship between the $2p$-th power and the p-th power derived above,

$$\mathbb{E}[Z^p] \leq \mathbb{E}[Z^{2p}]^{a/2p} \mathbb{E}[Z]^{p-a} \lesssim \mathbb{E}[Z^p]^{a/p}\, \mathbb{E}[Z]^{p-a}$$

so that

$$\mathbb{E}[Z^p] \lesssim \mathbb{E}[Z]^{\frac{p-a}{1-a/p}} = \mathbb{E}[Z]^p$$

Thus, we have that

$$\int_{\mathbb{R}^n} \|x\|_p \, d\gamma_n = \mathbb{E}[Z] \gtrsim \mathbb{E}[Z^p]^{1/p} \sim \sqrt{p}\, n^{1/p}$$

Lemma 31. *For $\infty > p > 2$, $D > 1$, if $l_2^k \overset{D}{\hookrightarrow} l_p^n$, then $k \lesssim p D^2\, n^{2/p}$. In other words, the result above is sharp for $p > 2$.*

Proof. Assume that $l_2^k \overset{D}{\hookrightarrow} l_p^n$. Then there exists a one to one linear map $T\colon l_2^k \to l_p^n$ such that $\|T\|\, \|T^{-1}\| \leq D$. By rescaling, we can assume without loss of generality that $\|T^{-1}\| = 1$, so that $\|T\| \leq D$. Let $e_1, \ldots, e_k$ be the standard basis of l_2^k. Write $Te_i = u_i = (u_{i1}, \ldots, u_{in}) \in \mathbb{R}^n$. Our assumption is that for all $a_1, \ldots, a_k \in \mathbb{R}$,

$$\left(\sum_{i=1}^k a_i^2\right)^{1/2} \leq \left(\sum_{j=1}^n \left|\sum_{i=1}^k a_i\, u_{ij}\right|^p\right)^{1/p} \leq D\left(\sum_{i=1}^k a_i^2\right)^{1/2}$$

where the left inequality follows from $\|T^{-1}\| \leq 1$ and the right inequality follows from $\|T\| \leq D$.

Now we play around with the inequalities. First fix $l \in \{1, \ldots, n\}$. We look at

$$\sum_{i=1}^k u_{il}^2 \leq \left(\sum_{j=1}^n \left|\sum_{i=1}^k u_{il}\, u_{ij}\right|^p\right)^{1/p} \leq D\left(\sum_{i=1}^k u_{il}^2\right)$$

where the first inequality follows since we are adding positive terms (when $j = l$ the left hand side comes out), and the second is from the right inequality earlier. This implies that

$$\left(\sum_{i=1}^k u_{il}^2\right)^{1/2} \leq D \quad \text{for all } 1 \leq l \leq n$$

Now we use the left ineqality above. For all $x_1, \ldots, x_n \in \mathbb{R}$, we have that

$$\left(\sum_{i=1}^k x_i^2\right)^{p/2} \leq \sum_{j=1}^n \left|\sum_{i=1}^k x_i\, u_{ij}\right|^p$$

(taking the left inequality and raising it to the p-th power).

Integrate both sides by $d\gamma_k$ and use Jensen's inequality:

$$\left(\int_{\mathbb{R}^k}\sum_{i=1}^k x_i^2\,d\gamma_k\right)^{p/2}\le\int_{\mathbb{R}^k}\left(\sum_{i=1}^k x_i^2\right)^{p/2}d\gamma_k\le\sum_{j=1}^n\int_{\mathbb{R}^k}\left|\sum_{i=1}^k x_i\,u_{ij}\right|^p d\gamma_k$$

Note that we can think of $\int_{\mathbb{R}^k}\left|\sum_{i=1}^k x_i\,u_{ij}\right|^p d\gamma_k$ as $\mathbb{E}_{g_1,\ldots,g_k}[|\sum g_i\,u_{ij}|^p]$ where g_1, ..., g_k are standard Gaussian random variables. We know that $\sum g_i\,u_{ij}$ also is a centered Gaussian distribution with variance $\sum u_{ij}^2$. Therefore

$$\begin{aligned}\sum_{j=1}^n\int_{\mathbb{R}^k}\left|\sum_{i=1}^k x_i\,u_{ij}\right|^p d\gamma_k &= \sum_{j=1}^n\left(\sum_{i=1}^k u_{ij}^2\right)^{p/2}\int_{\mathbb{R}}|x|^p d\gamma_1\\ &= n\,D^p\,(cp)^{p/2}\end{aligned}$$

The last equality is just the computation:

$$\begin{aligned}\int_{\mathbb{R}}|x|^p\,d\gamma_1 &= \frac{2}{\sqrt{2\pi}}\int_0^\infty x^p\,e^{-x^2/2}\,dx\\ &= \frac{2}{\sqrt{2\pi}}\int_0^\infty 2^{p/2}\,u^{\left(\frac{p+1}{2}-1\right)}e^{-u}\,du\\ &= \frac{2}{\sqrt{2\pi}}2^{p/2}\,\Gamma\left(\frac{p+1}{2}\right)\\ &\sim (Cp)^{p/2}\end{aligned}$$

using Stirling's approximation. □

Kashin's Volume Method

In this part we will investigate a different method for studying Euclidean sections. The same method will be used to prove the following theorems:

Theorem 32. (Kashin's First Theorem) *There exists $X, Y\subset\mathbb{R}^{2n}$ linear subspaces such that*

1. $\dim X=\dim Y=n$
2. *$X\perp Y$, i.e. $\langle x,y\rangle=0$ for all $x\in X, y\in Y$*
3. *For all $x\in X\cup Y$,*

$$\frac{1}{100}\sqrt{2n}\,\|x\|_2\le\|x\|_1\le\sqrt{2n}\,\|x\|_2$$

(Note the right inequality is always true by Cauchy Schwarz)

The third condition essentially says that $X\cup Y$ is essentially like l^1.

Theorem 33. (Kashin's Second Theorem) *There exists $U\in O(n)$ such that*

$$\frac{1}{\sqrt{n}}B_2^n\subset B_1^n\cap(UB_1^n)\subset\frac{100}{\sqrt{n}}\,B_2^n$$

This leads to the following theorem that we will not prove:

Theorem 34. (Litvak-Milman-Schechtman)

1. *For all* $\varepsilon > 0$*, there exists* $k(\varepsilon) \in \mathbb{N}$ *and* $U_1, \ldots, U_k \in O(n)$ *such that*

$$\frac{1}{\sqrt{n}} B_2^n \subset \bigcap_{i=1}^{k} U_1 B_1^n \subset \left(\frac{\pi}{2} + \varepsilon\right) \frac{1}{\sqrt{n}} B_2^n$$

2. *For all* $\varepsilon > 0$*, there exists* $C(\varepsilon) > 0$ *such that if* $U_1, \ldots, U_k \in O(n)$ *and*

$$\frac{1}{\sqrt{n}} B_2^n \subset \bigcap_{i=1}^{k} U_i B_1^n \subset \left(\frac{\pi}{2} - \varepsilon\right) \frac{1}{\sqrt{n}} B_1^n$$

then $k \geq e^{C(\varepsilon) n}$

(i.e. Kashin's method cannot find a subspace that goes down to $1 + \varepsilon$*)*

Theorem 35. (Kashin's Third Theorem) *For all* $\delta \in (0, 1)$*, there exists* $C(\delta) > 0$ *such that*

$$l_2^k \overset{C(\delta)}{\hookrightarrow} l_1^n, \; k \geq (1 - \delta) n$$

(Dvoretsky does not allow us to go arbitrarily close to n, just to some proportion of n)

Lemma 36. *Let* $K \subset \mathbb{R}^n$ *be the unit ball of some norm* $\|\cdot\|$*. Then*

$$\mathrm{vol}(K) = v_n \int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n} \, d\mu(x)$$

where $v_n = \mathrm{vol}(B_2^n) = \frac{\pi^n}{\Gamma(n/2 + 1)} \sim \frac{1}{n^{n/2}}$.

Proof. This is a computation in polar coordinates. For $f: \mathbb{R}^n \to \mathbb{R}^n$, we have

$$\int_{\mathbb{R}^n} f(x) \, dx = C_n \int_0^\infty \int_{\mathbb{S}^{n-1}} r^{n-1} f(r\theta) \, d\mu(\theta) \, dr$$

Note that if $f = \mathbf{1}_{B_2^n}$, we have that

$$\mathrm{vol}(B_2^n) = C_n \int_0^1 r^{n-1} dr = \frac{C_n}{n}$$

so that $C_n = n \, \mathrm{vol}(B_2^n)$.

Now for $f = \mathbf{1}_K$, we have that (using Fubini)

$$\begin{aligned} \mathrm{vol}(K) &= n v_n \int_{\mathbb{S}^{n-1}} \left(\int_0^{1/\|\theta\|} r^{n-1} \, dr \right) d\mu(\theta) \\ &= n v_n \int_{\mathbb{S}^{n-1}} \frac{1}{n \|\theta\|^n} \, d\mu(\theta) \\ &= v_n \int_{\mathbb{S}^{n-1}} \frac{1}{\|\theta\|^n} \, d\mu(\theta) \end{aligned}$$

$\square$

We will be making use of the following facts, to be proven as an exercise:

Exercise 5.

- $\mathrm{vol}(B_2^n)^{1/n} \approx \frac{1}{\sqrt{n}}$

  Just the polar computation using the Gaussian density $f(x) = \frac{1}{(2\pi)^{n/2}} e^{-\|x\|^2/2}$.

$$\begin{aligned}1 = \int_{\mathbb{R}^n} d\gamma_n &= n\,\mathrm{vol}(B_2^n) \int_0^\infty r^{n-1} \frac{1}{(2\pi)^{n/2}} e^{-r^2/2}\, dr \\ &= \frac{n}{(2\pi)^{n/2}} \mathrm{vol}(B_2^n) \int_0^\infty (2u)^{n/2-1} e^{-u}\, du \\ &= \frac{n}{2\pi^{n/2}} \mathrm{vol}(B_2^n) \Gamma\Big(\frac{n}{2}\Big) \\ &\sim \frac{n}{2\pi^{n/2}} \Big(\frac{n}{2e}\Big)^{n/2} \sqrt{\pi n}\ \mathrm{vol}(B_2^n) \\ &= \frac{\sqrt{\pi} n^{n/2+3/2}}{2(2\pi e)^{n/2}} \mathrm{vol}(B_2^n)\end{aligned}$$

- $\mathrm{vol}(B_\infty^n)^{1/n} = 2$

  Cube of side length 2.

- $\mathrm{vol}(B_1^n)^{1/n} = \Big(\frac{2^n}{n!}\Big)^{1/n} \approx \frac{1}{n}$

  Symmetry argument: $B_1^n = \{x\colon \sum |x_i| \leq 1\}$, so

$$\begin{aligned}\mathrm{vol}(B_1^n) &= \sum_{\varepsilon \in \{\pm 1\}^n} \mathrm{vol}\Big(\Big\{x\colon \mathrm{sgn}\, x_i = \varepsilon_i,\ \sum |x_i| \leq 1\Big\}\Big) \\ &= 2^n\, \mathrm{vol}\Big(\Big\{x\colon x_i \geq 0,\ \sum x_i \leq 1\Big\}\Big)\end{aligned}$$

  Now using the transformation

$$\Big\{x\colon x_i \geq 0,\ \sum x_i \leq 1\Big\} \xleftrightarrow{T} \{u\colon 0 \leq u_1 \leq \ldots \leq u_n \leq 1\}$$

  i.e. $x_i = u_i - u_{i-1}$, so $x = Tu$ with $T = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & \ddots & \\ & & & \ddots & -1 \\ & & & & 1 \end{pmatrix}$, since $\det(T) = 1$, we note that

$$\mathrm{vol}\Big\{x\colon x_i \geq 0,\ \sum x_i \leq 1\Big\} = \mathrm{vol}\, \{u\colon 0 \leq u_1 \leq \ldots \leq u_n \leq 1\}$$

  However, note that

$$\begin{aligned}1 = \mathrm{vol}\{u\colon 0 \leq u_i \leq 1\} &= \sum_\pi \mathrm{vol}\{u\colon 0 \leq u_{\pi(1)} \leq \ldots \leq u_{\pi(n)} \leq 1\} \\ &= n!\, \mathrm{vol}\{u\colon 0 \leq u_1 \leq \ldots \leq u_n \leq 1\}\end{aligned}$$

  we conclude that $\mathrm{vol}\{u\colon 0 \leq u_1 \leq \ldots \leq u_n \leq 1\} = \frac{1}{n!}$and thus

$$\mathrm{vol}(B_1^n) = \frac{2^n}{n!}$$

- For $0 < \delta \leq 2$, $x_0 \in \mathbb{S}^{n-1}$, if we look at the spherical cap

$$C_{\delta, x_0} := \{x \in \mathbb{S}^{n-1}\colon \|x - x_0\| \leq \delta\}$$

then $\mu(C_{\delta,x_0}) \geq \left(\frac{\delta}{2}\right)^n$.

Let's study the number of spherical caps needed to cover $\mathbb{S}^{n-1}$.

$$1 = \mu(\mathbb{S}^{n-1}) \leq (\#\,\text{caps})\,\mu(C_{\delta,x_0})$$

so that $\mu(C_{\delta,x_0}) \geq \frac{1}{\#\,\text{caps}}$. Note that the minimum number of caps needed is bounded above by the size of a maximal δ-separated set $\{x_i\}$, which we denote by m. Note that $B_\delta(x_i)$ forms a cover for $\mathbb{S}^{n-1}$, otherwise maximality is contradicted. We have that $\bigcup_i B_{\delta/2}(x_i) \subset B_{1+\delta/2}(0)$, where the union is disjoint, and by volume comparisons,

$$m\,(\delta/2)^n\,\mathrm{vol}(B_2^n) \leq (1+\delta/2)^n\,\mathrm{vol}(B_2^n)$$

and thus $m \leq (1+2/\delta)^n$ and $\mu(C_\delta, x_0) \geq (1+2/\delta)^{-n} = \left(\frac{\delta}{2+\delta}\right)^n$. Not sure how to improve this.

Theorem 37. *Let $K \subset \mathbb{R}^n$ be the unit ball of norm $\|\cdot\|$. Assume that*

1. $K \supset B_2^n$, *i.e.* $(\|x\| \leq \|x\|_2)$

2. $\left(\frac{\mathrm{vol}(K)}{\mathrm{vol}(B_2^n)}\right)^{1/n} \leq R$

Then there exists $U \in O(n)$ such that

$$B_2^n \subseteq K \cap (UK) \subseteq 8\,R^2 B_2^n$$

Remarks.

- Note that Kashin's second theorem follows immediately as a corollary. Using the facts above

$$\left(\frac{\mathrm{vol}(B_1^n)}{\mathrm{vol}\left(\frac{1}{\sqrt{n}}B_2^n\right)}\right)^{1/n} \sim \frac{2e}{\sqrt{2\pi e}} \leq \frac{5}{4}$$

- Also, Kashin's second thoerem implies Kashin's first theorem:

  If we have $U \in O(n)$ such that

$$\frac{1}{\sqrt{n}}B_2^n \subseteq B_1^n \cap (UB_1^n) \subseteq \frac{100}{\sqrt{n}}\,B_2^n$$

  let $X = \{(x, Ux), x \in \mathbb{R}^n\}$ and $Y = \{(-y, Uy), y \in \mathbb{R}^n\}$. Note that $X \perp Y$, since

$$\langle (x, Ux)\,, (-y, Uy)\rangle = -\langle x, y\rangle + \langle Ux, Uy\rangle = 0$$

  Now $B_1^n \cap UB_1^n = \{x \in \mathbb{R}^n$: max $\{\|x\|_1\,, \|Ux\|_1\} \leq 1\}$, and our assumption above can be rewritten as

$$\frac{1}{100}\sqrt{n}\,\|x\|_2 \leq \max\{\|x\|_1, \|Ux\|_1\} \leq \sqrt{n}\,\|x\|_2$$

  Now for $(x, Ux) \in X$, , we have that

$$\|(x, Ux)\|_{l_1^{2n}} = \|x\|_1 + \|Ux\|_1 \approx \max\{\|x\|_1,\, \|Ux\|_1\} \approx \sqrt{n}\,\|x\|_2 \approx \sqrt{n}\,\|(x, Ux)\|_{l_2^{2n}}$$

The same holds for $(-y, Uy)$ by the exact same argument.

Let us now prove the thoerem.

Proof. (of Theorem) For all $U \in O(n)$, define the norm $[x]_U = \frac{\|x\| + \|Ux\|}{2}$. By assumption we have

$$R^n \geq \frac{\mathrm{vol}(K)}{v_n} = \int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n} \, d\mu(x)$$

(the equality is from the polar coordinate computation of Lemma 36). We will show that there exists some $U \in O(n)$ such that

$$\int_{\mathbb{S}^{n-1}} \frac{1}{[x]_U} \, d\mu(x) \leq R^{2n}$$

Let ν be the Haar probability measure on $O(n)$. Then

$$\begin{aligned}
\int_{O(n)} \left(\int_{\mathbb{S}^{n-1}} \frac{1}{[x]_U^{2n}} \, d\mu(x) \right) d\nu(U) &= \int_{O(n)} \int_{\mathbb{S}^{n-1}} \frac{1}{\left(\frac{\|x\| + \|Ux\|}{2}\right)^{2n}} d\mu(x) \, d\nu(U) \\
\text{(AM-GM)} \quad &\leq \int_{O(n)} \int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n \|Ux\|^n} d\mu(x) \, d\nu(U) \\
&= \int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n} \left(\int_{O(n)} \frac{1}{\|Ux\|^n} \, d\nu(U) \right) d\mu(x) \\
&= \int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n} d\mu(x) \int_{\mathbb{S}^{n-1}} \frac{1}{\|y\|^n} \, d\mu(y) \\
&= \left(\int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n} d\mu(x) \right)^2 \leq R^{2n}
\end{aligned}$$

We have just shown that $\mathbb{E}_U\Big[\int_{\mathbb{S}^{n-1}} \frac{1}{[x]_U^{2n}} d\mu(x) \Big] \leq R^{2n}$ and thus there exists U for which this holds without the expectation.

Now, fix $y \in \mathbb{S}^{n-1}$, write $[y]_U = r$ with $r \leq 1$ (note $\|\cdot\| \leq \|\cdot\|_2$). Consider the spherical cap

$$C_r = \{x \in \mathbb{S}^{n-1}, \|x - y\|_2 \leq r\}$$

From the facts we know that $\mu(C_r) \geq \left(\frac{r}{2}\right)^n$.

Note

$$[x]_U \leq [x - y]_U + [y]_U \leq \|x - y\|_2 + [y]_U \leq 2r$$

Then we have that

$$\int_{\mathbb{S}^{n-1}} \frac{1}{[x]_U^{2n}} \, d\mu(x) \geq \int_{C_r} \frac{1}{(2r)^{2n}} \, d\mu(x) = \left(\frac{1}{2r}\right)^{2n} \left(\frac{r}{2}\right)^n = \left(\frac{1}{8r}\right)^n$$

On the other hand, we know that the integral on the left is bounded above by R^{2n}, thus $\frac{1}{8r} \leq R^2$ and

$$[y]_U = r \geq \frac{1}{8R^2}$$

Thus, for all $y \in \mathbb{R}^n$, we have that

$$[y]_U \geq \frac{1}{8R^2}\|y\|_2$$

and using $[x]_U \leq \max\{\|x\|, \|Ux\|\}$, we have

$$\max\{\|y\|, \|Uy\|\} \geq \frac{1}{8R^2}\|y\|_2$$

This implies the result, that

$$B_2^n \subseteq K \cap (UK) \subseteq 8R^2 B_2^n$$

□

We can adjust the method slightly to obtain Kashin's Third Theorem.

Theorem 38. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ with unit ball K such that*

1. *For all $x \in \mathbb{R}^n$, $\|x\| \leq \|x\|_2$*
2. $\left(\frac{\operatorname{vol} K}{\operatorname{vol}(B_2^n)}\right)^{1/n} \leq R$

Then for all $k \leq n$, there exists a subspace $V \subset \mathbb{R}^n$ with $\dim V = k$ and

$$\frac{1}{2^{\frac{n+k}{n-k}}} \cdot \frac{1}{R^{\frac{n}{n-k}}}\|x\|_2 \leq \|x\| \leq \|x\|_2 \text{ for all } x \in V$$

Remark: This proves the Third Theorem since we can take $k = (1-\delta)n$ and $K = B_1^n$, and we know from before that we can take $R = \frac{5}{4}$. Remember that while the Third Theorem allows k to be arbitrarily close to n, the embedding becomes worse (cannot get to $l_2^k \overset{1+\varepsilon'}{\hookrightarrow} l_1^n$).

Proof. As before we have that

$$\int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n}\, d\mu(x) = \frac{\operatorname{vol} K}{v_n} \leq R^n$$

Now denote $G_{n,k} = \{V \subset \mathbb{R}^n, V \text{ a subspace of } \dim k\}$. $O(n)$ acts on $G_{n,k}$ transitively, and with an appropriately defined metric $O(n)$ acts by isometries:

$$d_G(V, W) := \text{Hausdorff distance between } V \cap \mathbb{S}^{n-1} \text{and } W \cap \mathbb{S}^{n-1}$$

where the Hausdorff distance is the minimal d such that $W \cap \mathbb{S}^{n-1} \subset V \cap \mathbb{S}^{n-1} + dB_2^n$ and vice versa (how much we need to grow one set to swallow the other set). This is invariant under $O(n)$ by definition. Thus there is a unique Haar probability measure on $G_{n,k}$ which we denote $\nu_{n,k}$.

By uniqueness of Haar measure,

$$\int_{\mathbb{S}^{n-1}} \frac{1}{\|x\|^n}\, d\mu(x) = \int_{G_{n,k}} \int_{V \cap \mathbb{S}^{n-1}} \frac{1}{\|x\|^n}\, d\mu_V(x)\, d\nu_{n,k}$$

where μ_V is the measure μ restricted to $V\cap\mathbb{S}^{n-1}$ properly normalized. This says exactly that

$$\mathbb{E}_{G_{n,k}}\left[\int_{V\cap\mathbb{S}^{n-1}}\frac{1}{\|x\|^n}d\mu_V(x)\right]\le R^n$$

and thus there exists $V\in G_{n,k}$ for which $\int_{\mathbb{S}^{n-1}\cap V}\frac{1}{\|x\|^n}d\mu_V(x)\le R^n$.

We will then show that for all $x\in V\cap\mathbb{S}^{n-1}$, $\|x\|\ge\frac{1}{2^{\frac{n+k}{n-k}}}\cdot\frac{1}{R^{\frac{n}{n-k}}}$ using the spherical cap estimate as in the previous proof. Let $S^{k-1}=V\cap\mathbb{S}^{n-1}$. Fix $y\in S^{k-1}$, and denote $r=\|y\|\le 1$. Consider the cap

$$C_r=\{x\in S^{k-1}\colon\|x-y\|_2\le r\}$$

Again $x\in C_r$ implies that $\|x\|\le 2r$, and

$$R^n\ge\int_{S^{k-1}}\frac{1}{\|x\|^n}\,d\mu_V(x)\ge\int_{C_r}\frac{1}{(2r)^n}\,d\mu_V(x)=\frac{1}{(2r)^n}\mu_V(C_r)\ge\frac{1}{(2r)^n}\cdot\left(\frac{r}{2}\right)^k$$

Then simplifying gives the result, that

$$\|y\|=r\ge\frac{1}{2^{\frac{n+k}{n-k}}R^{\frac{n}{n-k}}}$$

for all $y\in S^{k-1}=V\cap\mathbb{S}^{n-1}$.

□

Week 6 (10/12/2010)

Summability in Banach Spaces

(Grothendieck)

Definition: A sequence $\{x_n\}_{n=1}^\infty$ in a Banach space X is

- unconditionally summable if $\sum_{n=1}^\infty x_{\sigma(n)}$ converges for all permutations σ of the integers
- absolutely summable if $\sum_{n=1}^\infty\|x_n\|<\infty$.

Remark: In every Banach space, absolutely summable sequences are unconditionally summable.

$$\left\|\sum_{n=m}^N x_{\sigma(n)}\right\|\le\sum_{n=m}^N\|x_{\sigma(n)}\|\longrightarrow 0\ \text{ as }\ m\to\infty\ (\min\{\sigma(n),n>m\}\longrightarrow\infty)$$

Dirichlet's Theorem: In $\mathbb{R}$ unconditional summability implies absolute summability. The same is also true for finite dimensional Banach spaces.

The following Lemma characterizes unconditional summability:

Lemma 39. *Let X be a Banach space and $\{x_n\}_{n=1}^\infty\subset X$. Then the following are equivalent:*

1. *$\{x_n\}_{n=1}^\infty$ is unconditionally summable*

2. *For all $\varepsilon > 0$, there exists $n_\varepsilon > 0$ such that for all finite index sets $S \subseteq \mathbb{N}$ with $\min S \geq n_\varepsilon$, we have that $\| \sum_{i \in S} x_i \| \leq \varepsilon$.*

3. *For all strictly increasing subsequences $\{x_{n_k}\}$, $\sum_{l=1}^{\infty} x_{n_k}$ converges.*

4. *For all $\varepsilon_1, \varepsilon_2, \ldots \in \{\pm 1\}$, $\sum_{i=1}^{\infty} \varepsilon_i x_i$ converges.*

Proof. $(1) \Longrightarrow (2)$. Suppose (2) does not hold, so that for some ε, we can find index sets J_i with $\max J_i < \min J_{i+1}$, and $\| \sum_{k \in J_i} x_k \| > \varepsilon$. Then we will find a permutation for which $\sum x_{\sigma(i)}$ is not summable, which is done by choosing σ so that J_i becomes contiguous $(= \sigma(\{a_i, a_i + 1, \ldots b_i - 1, b_i\}))$.

$(2) \Longrightarrow (1)$. Let σ, ε be given. By (2) we know there exists n_ε after which $\| \sum_{i \in S} x_i \| \leq \varepsilon$ for any finite S with $\min S \geq n_\varepsilon$. If we find $N_{\varepsilon,\sigma}$ for which $\{1, \ldots, n_\varepsilon\} \subset \{\sigma(1), \ldots, \sigma(N_{\varepsilon,\sigma})\}$, then it follows immediately that $\| \sum_m^n x_{\sigma(i)} \| \leq \varepsilon$ for $m, n > N_{\varepsilon,\sigma}$ ($i > N_{\varepsilon,\sigma} \Longrightarrow \sigma(i) > n_\varepsilon$).

$(2) \Longrightarrow (3)$ is also immediate, since we have convergence if for every $\varepsilon > 0$, we have $\| \sum_{k=m}^n x_{n_k} \| \leq \varepsilon$ for n, m sufficiently large, so we can take $S = \{n_k,\ n \leq k \leq m\}$.

$(3) \Longrightarrow (4)$ Let P be the index set for which $\varepsilon_i = 1$ and N be the index set for which $\varepsilon_i = -1$. Then by (3) we have that $\sum_{i \in P} x_i$ and $\sum_{i \in N} x_i$ both converge (ordered in increasing order). Now let $\varepsilon > 0$. Note that

$$\left\| \sum_{k=m}^{n} \varepsilon_i x_i \right\| \leq \left\| \sum_{\{m \leq k \leq n\} \cap P} x_i \right\| + \left\| \sum_{\{m \leq k \leq n\} \cap N} x_i \right\| \leq 2\varepsilon$$

for n, m sufficiently large.

$(4) \Longrightarrow (2)$. Suppose that (2) is false, then for some ε, we can find S_1, S_2, ... (each finite) with $\max S_i \leq \min S_{i+1}$ such that $\| \sum_{k \in S_i} x_k \| > \varepsilon$. Now set $\varepsilon_i = \begin{cases} 1 & i \in \bigcup_k S_k \\ -1 & \text{else} \end{cases}$. We then have

$$\sum_{i=1}^{\infty} (1 + \varepsilon_i) x_i = 2 \sum_{i \in \cup_k S_k} x_i$$

which diverges, and thus either $\sum_{i=1}^{\infty} x_i$ diverges or $\sum_{i=1}^{\infty} \varepsilon_i x_i$ diverges. Either way, we have found a sign sequence for which $\sum_{i=1}^{\infty} \varepsilon_i' x_i$ diverges, so (4) is false.

□

Remark 40. Here is an example that is unconditionally summable but not absolutely summable:

Take $X = l^\infty$, and e_i the standard basis. Take the sequence

$$e_1, \frac{1}{2} e_2, \frac{1}{2} e_3, \frac{1}{4} e_4, \ldots, \frac{1}{4} e_7, \ldots$$

i.e. definitely not absolutely summable. However, this sequence is summable:

For any ε, there exists n_0 after which $\| \sum_n^m c_i e_i \| = c_n \leq \varepsilon$. In fact, this is true for all choice of signs. Thus by the Lemma it is unconditionally summable.

It turns out that such examples can only come from infinite dimensional Banach spaces.

Theorem 41. (Dvoretsky-Rogers) *If a Banach space X has the property that unconditional summability implies absolute summability, then* $\dim X < \infty$.

Proof. We will prove the contrapositive. Assuming $\dim X = \infty$, the goal is to construct an unconditionally convergent sequence that is not absolutely convergent.

Take any $(a_i)_{i=1}^{\infty}$ with $\sum_{i=1}^{\infty} a_i^2 < \infty$. Find a subsequence $n_1 < n_2 < \ldots$ such that $\sum_{i \geq n_k} a_i^2 \leq 2^{-2k}$.

By the Dvoretsky-Rogers Lemma (Lemma 26), there exists $\{y_i\}_{i=1}^{\infty} \subset B_X$ such that $\|y_i\| \geq C$ (for some absolute constant C) and if $n_k \leq l < n_{k+1}$ and $\lambda_i \in \mathbb{R}$ then

$$\left\| \sum_{i=n_k}^{l} \lambda_i y_i \right\| \leq \sqrt{\sum_{i=n_k}^{l} \lambda_i^2}$$

(Take any arbitrarily large subspace of X and apply the Lemma) Now set $x_i = a_i \frac{y_i}{\|y_i\|}$ and let $\varepsilon_i = \pm 1$ be an arbitrary choice of signs. Then for $n_k \leq l \leq n_{k+1}$,

$$\left\| \sum_{i=n_k}^{l} \varepsilon_i x_i \right\| \leq \sqrt{\sum_{i=n_k}^{l} a_i^2 \cdot \frac{1}{\|y_i\|^2}} \leq \frac{1}{C} \sqrt{\sum_{i \geq n_k} a_i^2} \leq \frac{1}{C} 2^{-k}$$

and thus $\left\{ \sum_{i=1}^{n} \varepsilon_i x_i \right\}_n$ is a Cauchy sequence so that $\{x_i\}_{i=1}^{\infty}$ is unconditionally summable.

However, $\|x_i\| = a_i$ and we can easily choose a_i (i.e. $a_i = \frac{1}{i}$) for which a_i is not absolutely summable and that $\sum_{i=1}^{\infty} a_i^2 < \infty$. Thus $\{x_i\}_{i=1}^{\infty}$ is not absolutely summable. □

Lemma 42. *Let X, Y be Banach spaces, and let $T: X \to Y$ be linear. Then the following are equivalent:*

1. *For all unconditionally summable sequences $\{x_i\}_{i=1}^{\infty} \subset X$, $\{Tx_n\}$ is absolutely summable.*

2. *There exists $C > 0$ such that for all n and $x_1, \ldots, x_n \in X$, there exists a norm 1 linear functional $x^* \in X^*$ such that*

$$\sum_{i=1}^{n} \|Tx_i\| \leq C \sum_{i=1}^{n} |x^*(x_i)|$$

Proof. $(2) \Longrightarrow (1)$. Assume $\{x_n\}_{n=1}^{\infty}$ is unconditionally summable. Fix n. Let $x^* \in B_{X^*}$ be the functional from condition (2). Take $\varepsilon_i = \operatorname{sign}(x^*(x_i))$, then

$$\sum_{i=1}^{n} \|Tx_i\| \leq C x^* \left(\sum_{i=1}^{n} \varepsilon_i x_i \right) \leq C \left\| \sum_{i=1}^{n} \varepsilon_i x_i \right\|$$

and thus if x_i is unconditionally sumamble, Tx_i is absolutely summable.

$(1) \Longrightarrow (2)$. First, we prove the following claim: If $\{x_n\}_{n=1}^{\infty}$ is unconditionally summable, then

$$\sup_{x^* \in B_{X^*}} \sum_{i=1}^{\infty} |x^*(x_i)| < \infty$$

To show this, define $F: X^* \to l_1$ by $F(x^*) = (x^*(x_n))_{n=1}^{\infty} \in l^1$ (by unconditional convergence). Check that F has a closed graph $\{(x^*, Fx^*), x^* \in X^*\}$, i.e. if $(x_n^*, Fx_n^*) \to (x^*, y)$ then $y = Fx^*$.

$x_n^* \to x^*$ means that $x_n^*(x) \to x^*(x)$ for all x (in particular). $Fx_n^* \to y$ means that $\sum_i |x_n^*(x_i) - y_i| \to 0$. Now by Fatou's lemma,

$$\sum_i |x^*(x_i) - y_i| \leq \liminf_n \sum_i |x_n^*(x_i) - y_i| = 0$$

and thus $x^*(x_i) = y_i$, i.e. $y = Fx^*$

By the closed graph theorem this shows F is continuous, i.e.

$$\sum_{i=1}^{\infty} |x^*(x_i)| = \|F(x^*)\|_{l^1} \lesssim \|x^*\|_{X^*}$$

Taking the supremum over all $x^* \in B_{X^*}$ proves the claim.

Define Z to be the space of all unconditionally summable sequences in X. Define

$$\|(x_n)_{n=1}^{\infty}\| := \sup_{x^* \in B_{X^*}} \sum_i |x^*(x_i)|$$

Check that under this norm, Z is a Banach space.

To check that it is a norm, first we check that it is finite:

$$\sum_i |x^*(x_i)| = x^*\left(\sum_i \varepsilon_i x_i\right) \leq \left\|\sum_{i=1}^n \varepsilon_i x_i\right\| < \infty$$

Since the right hand side does not depend on x^*, taking the sup over $x^* \in B_{X^*}$ preserves finiteness.

Triangle inequality is immediate:

$$\|x_n + y_n\| \leq \sup_{x^* \in B_{X^*}} \sum_i |x^*(x_i)| + |x^*(y_i)| \leq \sup_{x^* \in B_{X^*}} \sum_i |x^*(x_i)| + \sup_{y^* \in B_{X^*}} \sum_i |y^*(y_i)| = \|x_n\| + \|y_n\|$$

and if $\|(x_n)_{n=1}^{\infty}\| = 0$, this means $\sum_i |x^*(x_i)| = 0$ for all x^* and thus $|x^*(x_i)| = 0$ for all i and all x^*. Then by duality we have that

$$\|x_i\| = \sup_{x^* \in B_{X^*}} |x^*(x_i)| = 0$$

and thus $x_i = 0$ for all i. Also need to check completeness...(?)

Let $l_1(Y)$ denote the space of all sequences y_i with the norm $\sum_{i=1}^{\infty} \|y_i\|$. Define $S: Z \to l_1(Y)$ by

$$S((x_i)) = (Tx_i)$$

Now we check that S has a closed graph.

This means checking that if $(Z_n, S(Z_n)) \to (\hat{Z}, \hat{Y})$ then $\hat{Y} = S(\hat{Z})$. Here $Z_n = (z_{n,1}, z_{n,2}, \ldots)$ and $\hat{Z} = (\hat{z}_1, , \hat{z}_2, \ldots)$ and $\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots)$. $Z_n \to \hat{Z}$ means that $z_{n,i} \to \hat{z}_i$ for all i (in particular). $S(Z_n) \to \hat{Y}$ means that $\sum_i \|Tz_{n,i} - \hat{y}_i\| \to 0$. Then again Fatou's lemma shows the result that $\sum_i \|T\hat{z}_i - \hat{y}_i\| = 0$.

This implies by the closed graph theorem that S is continuous, i.e.

$$\|(Tx_i)\|_{l_1(Y)} \lesssim \|(x_i)\|_Z = \sup_{x^* \in B_{X^*}} \sum_i |x^*(x_i)| < \infty$$

(where we have used the claim) In particular this implies the result.

□

p-Summability and Properties

Definition Let X, Y be Banach spaces, and let $1 \leq p < \infty$. We say that a linear map $T: X \to Y$ is **p-absolutely summable** if there exists $C > 0$ such that for all n and all $x_1, \ldots, x_n \in X$,

$$\left(\sum_{i=1}^{n} \|Tx_i\|^p \right)^{1/p} \leq C \sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^{n} |x^*(x_i)|^p \right)^{1/p}$$

The best constant C is denoted $\pi_p(T)$, the p-summing norm of T.

Facts.

- $\|T\| \leq \pi_p(T)$

  This follows from the case where $n = 1$, we note that $\|Tx\| \leq \pi_p(T) \sup_{x^*} |x^*(x)| = \pi_p(T) \|x\|$.

- The identity map $\mathrm{Id}: l_2 \to l_2$ is not 2-summing. Taking the standard basis $e_1, e_2, \ldots$ we have that

$$\sqrt{n} \leq C \sup_{\|x\|_2 \leq 1} \sqrt{\sum |\langle x, e_i \rangle|^2} \leq C$$

Lemma 43. *Let H_1, H_2 be Hilbert spaces. For all $T: H_1 \to H_2$,*

$$\pi_2(T) = \|T\|_{\mathrm{HS}} := \left(\sum_i \|Te_i\|^2 \right)^{1/2}$$

where e_i is an orthonormal basis of H_1 (can show that the HS norm does not depend on the choice of orthonormal basis).

For finite dimensions, then we can represent T as a matrix $T = (a_{ij})$ in which case the HS norm is the Frobenius norm $\sqrt{\sum_{ij} a_{ij}^2}$.

Proof. First, we note

$$\|T\|_{HS} = \left(\sum_{i=1}^{n} \|Te_i\|^2 \right)^{1/2} \leq \pi_2(T) \sup_{\|x\| \leq 1} \sqrt{\sum_{i=1}^{n} \langle x, e_i \rangle^2} = \pi_2(T)$$

We will write the opposite inequality in a slightly funny manner in anticipation of a later result. Let $\{g_j\}_{j \in J}$ be an orthonormal basis of H_2. Fix $x \in H_1$, then we have

$$\begin{aligned}
\|Tx\|^2 &= \sum_{j \in J} \langle Tx, g_j \rangle^2 = \sum_{j \in J} \langle x, T^* g_j \rangle^2 \\
&= \left(\sum_{j \in J} \|T^* g_j\|^2 \right) \sum_{j \in J} \left\langle x, \frac{T^* g_j}{\|T^* g_j\|} \right\rangle^2 \frac{\|T^* g_j\|^2}{\left(\sum_{j \in J} \|T^* g_j\|^2 \right)} \\
&= \|T\|_{HS}^2 \int_K (x^*(x))^2 \, d\mu(x^*)
\end{aligned}$$

where we note $\|T\|_{\mathrm{HS}}^2 = \|T^*\|_{\mathrm{HS}}^2 = \sum_j \|T^* g_j\|^2$ since

$$\sum_j \langle T^* g_j, T^* g_j\rangle = \sum_{i,j} \langle g_i, T^* g_j\rangle\langle T^* g_j, g_i\rangle = \sum_{i,j} \langle T g_i, g_j\rangle\langle g_j, T g_i\rangle = \sum_i \langle T g_i, T g_i\rangle$$

and where μ is just the atomic measure supported on the linear functionals $x \mapsto \left\langle x, \frac{T^* g_j}{\|T^* g_j\|}\right\rangle$ with weights $\|T^* g_j\|^2$ appropriately normalized. Thus we have shown that for all $x \in X$,

$$\|Tx\|^2 \leq \|T\|_{HS}^2 \int_K |x^*(x)|^2 \, d\mu(x^*)$$

and fixing $x_1, \ldots, x_n$, we have

$$\begin{aligned} \left(\sum \|Tx_i\|^2\right)^{1/2} &\leq \|T\|_{\mathrm{HS}} \sqrt{\int \left(\sum_{i=1}^n |x^*(x_i)|^2\right) d\mu(x^*)} \\ &\leq \|T\|_{\mathrm{HS}} \sup_{x^* \in B_{H_1^*}} \sqrt{\sum_{i=1}^n |x^*(x_i)|^2} \end{aligned}$$

so that $\pi_2(T) \leq \|T\|_{\mathrm{HS}}$. □

Lemma 44. (Ideal Property of π_p) *Suppose we have the maps*

$$U \xrightarrow{A} X \xrightarrow{T} Y \xrightarrow{B} V$$

then

$$\pi_p(BTA) \leq \|A\| \|B\| \pi_p(T)$$

Proof. For $x_1, \ldots, x_n \in X$, we have the computation

$$\begin{aligned} \left(\sum_{i=1}^n \|BTAx_i\|^p\right)^{1/p} &\leq \|B\| \left(\sum_{i=1}^n \|TAx_i\|^p\right)^{1/p} \\ &\leq \|B\| \pi_p(T) \sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n |x^*(Ax_i)|^p\right)^{1/p} \\ &\leq \|A\| \|B\| \pi_p(T) \sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n \left|\frac{x^* \circ A}{\|A\|}(x_i)\right|^p\right)^{1/p} \\ &\leq \|A\| \|B\| \pi_p(T) \sup_{u^* \in B_{U^*}} \left(\sum_{i=1}^n |u^*(x_i)|^p\right)^{1/p} \end{aligned}$$

□

The p-summability operators follows a similar inclusion as for l^p spaces, in that it is easier to be summable for higher powers:

Lemma 45. *Let $1 \leq r < p < \infty$ and $T: X \to Y$. Then*

$$\pi_p(T) \leq \pi_r(T)$$

Proof. Let q satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, i.e. $\frac{1}{p/r} + \frac{1}{q/r} = 1$. Take $x_1, \ldots, x_n \in X$ such that

$$\sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} = 1$$

By homogeneity it suffices to show that

$$\left(\sum_{i=1}^n \|Tx_i\|^p \right)^{1/p} \leq \pi_r(T)$$

Let $(c_1, \ldots, c_n)$ such that $\sum_{i=1}^n |c_i|^q = 1$. Then using Hölder,

$$\left(\sum_{i=1}^n |c_i|^r \, |x^*(x_i)|^r \right)^{1/r} \leq \left(\sum_{i=1}^n |c_i|^q \right)^{1/q} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} = \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p}$$

By definition of $\pi_r(T)$, we have

$$\begin{aligned} \left(\sum_{i=1}^n |c_i|^r \, \|Tx_i\|^r \right)^{1/r} &\leq \pi_r(T) \sup_{x^*} \left(\sum_{i=1}^n |c_i|^r \, |x^*(x_i)|^r \right)^{1/r} \\ &\leq \pi_r(T) \sup_{x^*} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} \\ &= \pi_r(T) \end{aligned}$$

Now if we maximize the left hand side subject to the constraint $\|c\|_q = 1$, will get that $c_i = \frac{\|Tx_i\|^{p/r-1}}{\left(\sum \|Tx_i\|^{q(p/r-1)} \right)^{1/q}}$ and plugging this value of c_i above gives the result.

$$|c_i|^r = \frac{\|Tx_i\|^{p-r}}{\left(\sum \|Tx_i\|^{q(p/r-1)} \right)^{r/q}}$$

$$|c_i|^r \, \|Tx_i\|^r = \frac{\|Tx_i\|^p}{\left(\sum \|Tx_i\|^{q(p/r-1)} \right)^{r/q}}$$

$$\left(\sum |c_i|^r \, \|Tx_i\|^r \right)^{1/r} = \frac{(\sum \|Tx_i\|^p)^{1/r}}{(\sum \|Tx_i\|^p)^{1/q}} = \left(\sum_{i=1}^n \|Tx_i\|^p \right)^{1/p}$$

(Note: $p + q = qp/r$ and $1/r - 1/q = 1/p$)

$\square$

Proposition 46. *Assume $T: X \to Y$ with $p \geq 1$, and $K \subseteq B_{X^*}$ is a norming subset, i.e. for all $x \in X$,*

$\|x\| = \sup_{x^* \in K} |x^*(x)|$ *(i.e. can use K in the dual expression for norm instead of all of B_{X^*}). Then*

$$\sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} = \sup_{x^* \in K} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} = \sup_{\|\alpha\|_{p^*}=1} \left\| \sum_{i=1}^n \alpha_i x_i \right\|$$

Proof. Let $p^* = \frac{p}{p-1}$. Then by duality,

$$\begin{aligned} \sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} &= \sup_{x^* \in B_{X^*}} \sup_{\|\alpha\|_{p^*}=1} \sum_{i=1}^n x^*(x_i) \alpha_i \\ &= \sup_{\|\alpha\|_{p^*}=1} \sup_{x^* \in B_{X^*}} x^* \left(\sum_{i=1}^n \alpha_i x_i \right) \\ &= \sup_{\|\alpha\|_{p^*=1}} \left\| \sum_{i=1}^n \alpha_i x_i \right\| \end{aligned}$$

where $\alpha \in (\mathbb{R}^n, L^{p^*})$ above. Now the same computation holds if we replace X^* by K since K is assumed to be a norming subset. This implies the result. ☐

Lemma 47. *Let $T: X \to Y$. Then*

$$\begin{aligned} \pi_p(T) &= \sup \{ \pi_p(TS) \ \ s.t.\ \ S: l_{p^*}^n \to X, \|S\| \le 1 \} \\ &= \sup \left\{ \left(\sum_{i=1}^n \|TSe_i\|^p \right)^{1/p} \ \ s.t.\ \ S: l_{p^*}^n \to X, \|S\| \le 1 \right\} \end{aligned}$$

Proof. Denote $K = \sup \{ \pi_p(TS) \text{ s.t. } S: l_{p^*}^n \to X, \|S\| \le 1 \}$. Note that the ideal property shows that $K \le \pi_p(T)$ ($\pi_p(TS) \le \|S\| \pi_p(T)$). Choose $x_1, \ldots, x_n$ such that both

$$\begin{aligned} \sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} &= 1 \\ \left(\sum_{i=1}^n \|Tx_i\|^p \right)^{1/p} &\ge \pi_p(T) - \varepsilon \end{aligned}$$

(definition of $\pi_p(T)$ being the sharpest constant for which the inequality holds). Now define $S: l_{p^*}^n \to X$ by

$$S((\alpha_i)_{i=1}^n) = \sum_{i=1}^n \alpha_i x_i$$

From the computation in the previous fact, we note that $\|S\| = 1$ since

$$\sup_{x^* \in B_{X^*}} \left(\sum_{i=1}^n |x^*(x_i)|^p \right)^{1/p} = \sup_{\|\alpha\|_{p^*} \le 1} \sum_{i=1}^n \alpha_i x_i = 1$$

Then we have

$$\left(\sum_{i=1}^n \|TSe_i\|^p \right)^{1/p} = \left(\sum_{i=1}^n \|Tx_i\|^p \right)^{1/p} \ge \pi_p(T) - \varepsilon$$

on one hand, and

$$\left(\sum_{i=1}^n \|TSe_i\|^p\right)^{1/p} \le \pi_p(TS) \sup_{\|\beta\|_p\le 1}\left(\sum \langle\beta, e_i\rangle^p\right)^{1/p} = \pi_p(TS)$$

(note here that TS is an operator on $l_{p^*}^n$, so the dual is l_p^n). Combining the last two observations we have that

$$\pi_p(T) \le \pi_p(TS) + \varepsilon \le \sup_S \pi_p(TS) + \varepsilon$$

and as ε is arbitrary, we have the result. □

Week 7 (10/19/2010)

Here is another observation about p-summing operators.

Proposition 48. *Let μ be any probability measure on Ω. Then the identity operator* Id: $L_\infty(\mu) \to L_p(\mu)$ *satisfies* $\pi_p(\mathrm{Id}) \le 1$

Proof. For all $a = (a_1, ..., a_n) \in l_{p^*}^n$, and $f_1, ..., f_n \in L_\infty(\mu) \subset L_p(\mu)$ there exists $E_a \subset \Omega$ of measure zero such that for all $\omega \in E_a$,

$$\left|\sum_{i=1}^n a_i f_i(\omega)\right| \le \left\|\sum_{i=1}^n a_i f_i\right\|_\infty$$

Let $\mathcal{F} \subset l_{p^*}^n$ be a countable dense subset, then $E = \bigcup_{a\in\mathcal{F}} E_a$ has measure zero. Then

$$\begin{aligned}\left(\sum_{i=1}^n \|f_i\|_{L^p(\mu)}^p\right)^{1/p} &= \left(\int_\Omega \left(\sum_{i=1}^n |f_i(\omega)|^p\right) d\mu(\omega)\right)^{1/p} \\ &\le \sup_{\omega\in\Omega\setminus E}\left(\sum_{i=1}^n |f_i(\omega)|^p\right)^{1/p} \\ &= \sup_{\omega\in\Omega\setminus E}\ \sup_{\substack{a\in\mathcal{F}\\ \|a\|_{p^*}\le 1}}\left|\sum_{i=1}^n a_i f_i(\omega)\right| \\ &\le \sup_{a\in l_{p^*}^n}\left\|\sum_{i=1}^n a_i f_i\right\|_\infty\end{aligned}$$

which shows using the definition and Proposition 46 that $\pi_p(\mathrm{Id}) \le 1$.

□

Pietsch Domination Theorem

Theorem 49. (Pietsch Domination) *Let X, Y Banach spaces, $p \ge 1$, and let $K \subseteq B_{X^*}$ norming and weak*-closed.*

If $T: X \to Y$ is p-summing then there exists regular Borel Prob. measure μ on K such that for all $x \in X$,

$$\|Tx\| \le \sigma_p(T)\left(\int_K |x^*(x)|^p\, d\mu(x^*)\right)^{1/p}$$

Exercise 6. Show that if an operator T satisfies the above inequality for some Borel Probability measure that T is p-summing.

Solution:

$$\begin{aligned}\Big(\sum \|Tx_i\|^p\Big)^{1/p} &\le \sigma_p(T)\Big(\int_K \sum |x^*(x_i)|^p\, d\mu(x^*)\Big)^{1/p}\\ &\le \sigma_p(T) \sup_{x^*\in K} \Big(\sum |x^*(x_i)|^p\Big)^{1/p}\\ &= \sigma_p(T) \sup_{x^*\in B_{X^*}} \Big(\sum |x^*(x_i)|^p\Big)^{1/p}\end{aligned}$$

Interpretation:

For a fixed $x \in X$, the right side of the inequality above is the $L^p(\mu)$ norm of the function $f_x(x^*) = x^*(x)$ for $x^* \in K$, i.e. $f_x \in C(K) \hookrightarrow L_p(K, \mu)$.

Define $J: X \to C(K)$ by $J(x) = f_x$, i.e. $J(x)(x^*) = x^*(x)$. Then

$$\|J(x)\|_\infty = \sup_{x^*\in K} |x^*(x)| = \sup_{x^*\in B_{X^*}} |x^*(x)| = \|x\|$$

using the fact that K is norming. This shows that J is an isometry, and is invertible as a map from $X \to JX$.

Denote the I by the identification of $C(K)$ as elements of $L_p(K, \mu)$. Note

$$\|IJ(x)\|_{L^p(\mu)} = \Big(\int_K |x^*(x)|^p\, d\mu(x^*)\Big)^{1/p} \le \|J(x)\|_\infty$$

so that $\|I\|_{C(K)\to L^p(\mu)} \le 1$, and let $X_p = IJX$ be the range of I on JX. In particular, I is invertible as a map from $JX \to X_p$ by definition.

Then we have the following diagram:

$$\begin{array}{ccc} C(K) & \xrightarrow{I} & L_p(K,\mu)\\ \cup & & \cup\\ JX & \xrightarrow[I|_{JX}]{} & X_p\\ J\Big\uparrow & & \Big\downarrow S\\ X & \xrightarrow{T} & Y \end{array}$$

and we can define S so that the diagram above commutes, i.e. $S(IJx) = Tx$ (I, J are invertible).

We note that as an operator $S: X_p \to Y$, the norm is

$$\|S(IJx)\|_Y^p = \|Tx\|_Y^p \le \pi_p(T)^p \int_K |x^*(x)|^p d\mu(x^*) = \pi_p(T)^p\, \|IJx\|_{L^p(\mu)}^p$$

and thus $\|S\|_{X_p\to Y} \le \pi_p(T)$.

Special case $p = 2$. In this case, we have that $X_p = L_2(\mu)$ is a Hilbert space, and we can use the orthogonal projection from $L_2(\mu) \to \overline{X_2}$. Then let $A = \mathrm{Proj} \circ I \circ J$ and $B = S$, again $\|B\| \le \pi_2(T)$ from the above, and from the ideal property (Lemma 44),

$$\pi_2(A) \le \|\mathrm{Proj}\|\, \|J\|\, \pi_2(I) \le 1$$

so that $\pi_2(A) \le 1$. Above we used Proposition 48, where we showed that $I: L^\infty \to L^p$ is p-summing with norm 1. Thus we have shown the following useful corollary:

Corollary 50. (Pietsch Factorization) *If $T: X \to Y$ is 2-summing, then there exists a Hilbert space H, and mappings $A: X \to H$, $B: H \to Y$ such that the following diagram commutes:*

$$\begin{tikzcd} & H \arrow[dr, "B"] & \\ X \arrow[ur, "A"] \arrow[rr, "T"] & & Y \end{tikzcd}$$

i.e. $T = BA$ and $\|B\| \le \pi_2(T)$, $\pi_2(A) \le 1$.

From here we can prove an interesting result:

Theorem 51. *Let X be an n-dimensional space, $I_X: X \to X$ be the identity map. Then*

$$\pi_2(I_X) = \sqrt{n}$$

Proof. By Pietsch Factorization (Corollary 50), we have a Hilbert space H, and mappings A, B so that the following diagram commutes:

$$\begin{tikzcd} & H \arrow[dr, "B"] & \\ X \arrow[ur, "A"] \arrow[rr, "I_X"] & & X \end{tikzcd}$$

and that $\pi_2(A) \le 1$, $\|B\| \le \pi_2(I_X)$. Replacing H with AX we can assume without loss of generality that A is onto. We have $I_X = BA$, and it must be the case that $\dim H = n$ (otherwise impossible to get identity), and since we are working in finite dimensions, $B = A^{-1}$. Then denoting the identity on H by I_H, we have that since $I_X = B I_H A$, $I_H = B^{-1} I_X A^{-1} = AB$. This implies that

$$\sqrt{n} = \pi_2(I_H) \le \pi_2(A)\,\|B\| \le \pi_2(I_X)$$

where we have used the ideal property of π_2 (Lemma 44) above and the fact that π_2 in a Hilbert space is just the Hilbert Schmidt norm (Lemma 43).

To show the upper bound, we use Lemma 47:

$$\pi_2(I_X) = \sup\{\pi_2(I_X S),\ S: l_2^m \to X, \|S\| \le 1, m \in \mathbb{N}\}$$

For $\varepsilon > 0$ let S be chosen such that $\pi_2(I_X) \le \pi_2(S) - \varepsilon$. We will use the isomorphism $H := l_2^m/\ker(S) \cong S(l_2^m)$. Define $Q: l_2^m \to H$ to be the natural identification of l_2^m to the quotient $l_2^m/\ker(S)$, i.e. $Qx = x + \ker(S) \in H$, which is an isometry ($\|Q\| \le 1$), and define $\tilde{S}$ so that $S = Q I_H \tilde{S}$. In particular, $\|\tilde{S}\| = \|S\| \le 1$ and by the ideal property,

$$\pi_2(I_X) \le \pi_2(S) - \varepsilon = \pi_2(Q I_H \tilde{S}) - \varepsilon \le \|Q\|\,\pi_2(I_H)\,\|\tilde{S}\| - \varepsilon = \sqrt{n} - \varepsilon$$

and since ε is arbitrary, we have the upper bound. □

Corollary 52. *Let $K \subset \mathbb{R}^n$ be a centrally symmetric convex body. Then there exists an ellipsoid $\mathcal{E} \subset K$ such that $\sqrt{n}\,\mathcal{E} \supset K$. (Later we will see that John's ellipsoid satisfies this inequality).*

Exercise 7. Show that the $\sqrt{n}$ is sharp above, looking at $K = [-1,1]^n$.

Proof. Let X be the Banach space whose unit ball is K. Denote I_X by the identity map on X. By Pietsch Factorization (Corollary 50) again, we have a Hilbert space H, and mappings A, B so that the following diagram commutes:

$$\begin{array}{ccc} & H & \\ {}^{A}\nearrow & & \searrow_{B} \\ X & \xrightarrow{I_X} & X \end{array}$$

and $\pi_2(A) \leq 1$, $\|B\| \leq \pi_2(I_X) = \sqrt{n}$ (using the previous result). As before, we may assume without loss of generality that $H = AX$, and again we have dim $H = n$ and $BA = I_X$ so that $B = A^{-1}$. Now choose $\mathcal{E} = \frac{A^{-1}(B_2^n)}{\sqrt{n}}$. Now we see that if $x \in \mathcal{E} = \frac{1}{\sqrt{n}} A^{-1}(B_2^n)$, this means that $\|Ax\|_2 \leq \frac{1}{\sqrt{n}}$. But then we have that

$$\|x\| = \|BAx\| \leq \|B\| \, \|Ax\|_2 \leq 1$$

so that $x \in K$, i.e. $\mathcal{E} \subset K$. Then for $x \in K$, we have that $\|x\| \leq 1$ and $\|Ax\|_2 \leq \|A\| \, \|x\| \leq \pi_2(A) \leq 1$ so that $x \in A^{-1}(B_2^n) = \sqrt{n}\, \mathcal{E}$.

$\square$

Now we turn to the proof of the Pietsch Domination Theorem.

Proof. (of Pietsch Domination, Theorem 49) By the Theorem of Banach-Alaoglu, we know that K (the norming subset of B_{X^*} that is weak* closed) is compact in the weak* topology. For all $x_1, \ldots, x_n \in X$, define $g_{(x_1, \ldots, x_n)}\colon K \to \mathbb{R}$ by

$$g_{(x_1, \ldots, x_n)}(x^*) = \sum_{i=1}^{n} \|Tx_i\|^p - \pi_p(T)^p \sum_{i=1}^{n} |x^*(x)|^p$$

Let $C(K) = (C(K), \|\cdot\|_{\sup})$ be the continuous functions on K with the sup norm.

$g_{(x_1, \ldots, x_n)} \in C(K)$. Define $Q \subset C(K)$ by

$$Q = \left\{ g_{(x_1, \ldots, x_n)}\colon n \in \mathbb{N},\, x_1, \ldots, x_n \in X \right\}$$

We check that Q is a convex set since for $\lambda \in [0, 1]$ and $x_1 \ldots, x_n, y_1, \ldots, y_m \in X$, we have that

$$\lambda g_{(x_1, \ldots, x_n)} + (1 - \lambda)\, g_{(y_1, \ldots, y_m)} = g_{\left(\lambda^{1/p} x_1, \ldots, \lambda^{1/p} x_n, (1-\lambda)^{1/p} y_1, \ldots, (1-\lambda)^{1/p} y_m\right)}$$

Now let $P = \{f \in C(K)\colon f(x^*) > 0 \text{ for all } x^* \in K\}$. P is easily convex, and it is also open since K is compact ($f \in P$ achieves its minimum on K, say it is ε, then $\{g, \|g - f\|_\infty < \varepsilon\}$ is an open set contained in P).

Note that $P \cap Q = \{\}$, otherwise there exists $x_1, \ldots, x_n \in X$ such that

$$\sum_{i=1}^{n} \|Tx_i\|^p > \pi_p(T)^p \sum_{i=1}^{n} |x^*(x_i)|^p \quad \text{for all } x^* \in K$$

which contradicts T being p-summing with constant $\pi_p(T)$.

By the hyperplane separation theorem (geometric Hahn Banach) and the Riesz Representation theorem for $C(K)$, there exists a regular Borel measure μ on K and $c \in \mathbb{R}$ such that for all $g \in Q$ and $f \in P$,

$$\int_K g\, d\mu \leq c < \int_K f\, d\mu$$

since $0 \in Q$, c must be nonnegative, and thus we have that for all $f \geq 0$ (i.e. in P), $\int_K f d\mu \geq 0$ so that μ is a positive measure. For all $\varepsilon > 0$, if we take $\varepsilon \mathbf{1}_K \in P$ we have that

$$c < \varepsilon \mu(K)$$

and since ε is arbitrary this implies $c = 0$. By normalizing, without loss of generality μ is a probability measure. For $x \in X$, apply $\int_K g d\mu \leq 0$ to $g = g_{\{x\}}$, then

$$\int_K (\|Tx\|^p - \pi_p(T)^p |x^*(x)|^p) \, d\mu(x^*) \leq 0$$

and this implies the existence of some x^* for which

$$\|Tx\|^p \leq \pi_p(T)^p \int_K |x^*(x)|^p \, d\mu(x^*)$$

which proves the result. □

Notation: Let X, Y be Banach spaces, and denote

$$\begin{aligned} \Pi_p(X,Y) &= \{T: X \to Y, \pi_p(T) < \infty\} \\ \mathcal{L}(X,Y) &= \{\text{bounded operators from } X \to Y\} \end{aligned}$$

Recall that for $1 \leq r < p < \infty$, that $\Pi_r(X,Y) \subset \Pi_p(X,Y)$, i.e. for $T: X \to Y$, $\pi_p(T) \leq \pi_p(T)$ (Lemma 45)

Next, we show that if the reverse inequality is true up to a constant, then we can show the same for the 1-summing norm.

Theorem 53. (Maurey Extrapolation Theorem) *Let X be a Banach space, and $1 \leq r < p < \infty$ such that*

$$\Pi_p(X, l_p) = \Pi_r(X, l_p)$$

Then for all Banach spaces Y,

$$\Pi_1(X,Y) = \Pi_p(X,Y)$$

Remarks

1. Using the closed graph theorem, the hypothesis $\Pi_p(X, l_p) = \Pi_r(X, l_p)$ is the same as the following:

   There exists a constant $C > 0$ such that for all $T: X \to l_p$,

   $$\pi_r(T) \leq C \pi_p(T)$$

   Let $\Pi = \Pi_p = \Pi_r$. That this condition is sufficient follows immediately. To show the necessity, we need that the identity map $\mathrm{Id}: (\Pi, \pi_p) \to (\Pi, \pi_r)$ is continuous (which implies the inequality). Closed graph theorem says that if the graph $(T, T) \in \Pi_p \times \Pi_r$ is closed, then T is continuous. In other words, given a sequence T_k for which $T_k \to T^{(p)}$ in Π_p and $T_k \to T^{(r)}$ in Π_r that $T^{(p)} = T^{(r)}$. Note that since $\|T\| \leq \min\,(\pi_p(T), \pi_r(T))$, we know that $\|T_k - T^{(p)}\| \to 0$ and $\|T_k \to T^{(r)}\| \to 0$. By uniqueness of limits for $\|\cdot\|$ we have that $T^{(p)} = T^{(r)}$ as desired.

Quantitative Version: We will end up showing that for all $T: X \to Y$,

$$\pi_1(T) \leq 2(2c)^{\frac{r(p-1)}{p-r}} \pi_p(T)$$

2. We note that the condition $\pi_r(T) \leq C\pi_p(T)$ for all $T: X \to l_p$ implies that

$$\pi_r(T) \leq C\pi_p(T) \text{ for all } T: X \to l_p^n$$

and this statement in turn implies that for $\varepsilon > 0$,

$$\pi_r(T) \leq C(1+\varepsilon)\pi_p(T) \text{ for all } T: X \to L^p(\mu)$$

Proof. (From $l_p \to l_p^n$) Given $T: X \to l_p^n$, we can extend trivially to a function $\tilde{T}: X \to l_p$ where $\tilde{T}x = (\vdash Tx \dashv, 0, 0, \ldots)$, and we note that

$\pi_p(T) = \pi_p(\tilde{T})$, as $\|Tx\|_{l_p^n} = \|Tx\|_{l_p}$ for all x. So by assumption since $\pi_r(\tilde{T}) \leq C\pi_p(\tilde{T})$, we have immediately that $\pi_r(T) \leq C\pi_p(T)$. □

Proof. (From $l_p^n \to L_p(\mu)$) Let $T: X \to L_p(\mu)$, and fix $x_1, \ldots, x_n \in X$. Let $E = \operatorname{span}\{Tx_1, \ldots, Tx_n\}$.

Let $f_1, \ldots, f_n \in B_X$ be a basis for E. Because all norms are equivalent in finite dimensions, we note that there is some constant $M > 0$ with

$$\sum_{i=1}^n |a_i| \leq \left\| \sum_{i=1}^n a_i f_i \right\|_{L_p(\mu)} \leq M \sum_{i=1}^n |a_i|$$

Let $h_1, \ldots, h_n \in L_p(\mu)$ be simple functions such that $\|f_i - h_i\|_{L_p(\mu)} \leq \frac{\varepsilon}{M}$. Let $A_1, \ldots, A_N$ be the subsets of positive measure for which all the h_i are constant.

Let $F = \operatorname{span}\{\mathbf{1}_{A_i}\}_{i=1}^N$. If we define $S: F \to l_p^N$ by $S(\mathbf{1}_{A_i}) = \mu(A_i)^{1/p} e_i$, then S is an isometry.

Now define $U: E \to F$ by the mapping $Uf_i = h_i$. Note that

$$\begin{aligned}
\left\| \sum_{i=1}^N a_i f_i - \sum_{i=1}^N a_i h_i \right\|_{L_p} &\leq M \sum_{i=1}^N |a_i| \, \|f_i - h_i\|_{L_p} \\
&\leq M \sum_{i=1}^N |a_i| \frac{\varepsilon}{M} \\
&\leq \left\| \sum_{i=1}^N a_i f_i \right\|_{L_p} \varepsilon
\end{aligned}$$

or in other words, using triangle inequality we have that

$$(1-\varepsilon) \left\| \sum_{i=1}^N a_i f_i \right\|_{L_p} \leq \left\| \sum_{i=1}^N a_i h_i \right\|_{L_p} \leq (1+\varepsilon) \left\| \sum_{i=1}^N a_i f_i \right\|_{L_p}$$

so that $\|U\| \leq 1 + \varepsilon$ and $\|U^{-1}\| \leq \frac{1}{1-\varepsilon}$. The rest follows from the ideal property:

Let P be the projection from $L_p(\mu) \to F$ defined by the conditional expectation:

$$P(f) = \mathbb{E}[f \mid \sigma(h_1, \ldots, h_N)]$$

noting that $\|P\| \leq 1$ and that $P(L_p(\mu)) = \operatorname{span}\{\mathbf{1}_{A_i}\} = F$ (note that the algebra generated by the indicators is exactly the linear span since $\mathbf{1}_{A_i}^k = \mathbf{1}_{A_i}$).

Now applying our result, we know that since $SPT: X \to l_p^N$, the ideal property tells us that

$$\pi_r(SPT) \leq C\pi_p(SPT) \leq C\pi_p(T)$$

This means that plugging in UTx_i, and noting that $P(UTx_i) = UTx_i$ since $UTx_i \in F$, we have that

$$\Big(\sum \|UTx_i\|^r\Big)^{1/r} = \Big(\sum \|SPUTx_i\|^r\Big)^{1/r} \leq C(1+\varepsilon)\pi_p(T)\sup_{x^*}\Big(\sum |x^*(x_i)|^r\Big)^{1/r}$$

and using the fact that $\|Tx_i\| \leq \frac{1}{1-\varepsilon}\|UTx_i\|$, we conclude that

$$\Big(\sum \|Tx_i\|^r\Big)^{1/r} \leq \frac{C}{1-\varepsilon}\pi_p(T)\sup_{x^*}\Big(\sum |x^*(x_i)|^r\Big)^{1/r}$$

so that $\pi_r(T) \leq \frac{C}{1-\varepsilon}\pi_p(T)$. □

Proof. (of Maurey Extrapolation, Theorem 53) By assumption, there exists C such that for all measures μ, and for all $T: X \to L_p(\mu)$, we have that $\pi_r(T) \leq C\pi_p(T)$. Denote $K = B_{X^*}$ with weak* topology, and denote $P(K)$ to be the set of all regular Borel probability measures on K.

Let $J: X \to C(K)$ be the duality mapping $J(x)(x^*) = x^*(x)$ as in the interpretation of Pietsch, let I_μ be the identity map from $C(K) \to L_p(K, \mu)$ and let $J_\mu = I_\mu J: X \to L_p(K, \mu)$, defined for each $\mu \in P(K)$. Note that $\pi_p(J_\mu) \leq 1$ by the ideal property since $\pi_p(I_\mu) \leq 1$ and $\|J\| \leq 1$. By assumption, we have $\pi_r(J_\mu) \leq C$. Now by Pietsch, there exists $\hat{\mu} \in P(K)$ such that for all $x \in X$,

$$\|J_\mu x\|_{L_p(\mu)} \leq C\,\|J_{\hat{\mu}} x\|_{L_r(\hat{\mu})}$$

Now take $T: X \to Y$ with $\pi_p(T) < \infty$. We want to show that T is 1-summing. By Pietsch there exists $\mu_0 \in P(K)$ such that for all $x \in X$

$$\|Tx\| \leq \pi_p(T)\,\|J_{\mu_0}(x)\|_{L_p(\mu_0)}$$

We claim that there exists $\lambda \in P(K)$ such that for all $x \in X$

$$\|J_{\mu_0} x\|_{L_p(\mu_0)} \leq C'\,\|J_\lambda x\|_{L^1(\lambda)}$$

which shows that $\pi_1(T) \leq C'\pi_p(T)$, since in this case for $x_1, \ldots, x_n \in X$,

$$\sum_{i=1}^n \|Tx_i\| \leq \pi_p(T) C' \sum_{i=1}^n \int |x^*(x_i)|\, d\lambda(x^*) \leq C'\pi_p(T) \sup_{x^* \in B_{X^*}} \sum_{i=1}^n |x^*(x_i)|$$

Now define $\mu_{n+1} = \widehat{\mu_n}$ from Pietsch, i.e.

$$\|J_{\mu_n} x\|_{L_p(\mu)} \leq C\,\|J_{\mu_{n+1}} x\|_{L_r(\mu_{n+1})}$$

(as above) and set $\lambda = \sum_{n=0}^\infty 2^{-n-1}\mu_n$. Then we show that λ works. Let $\frac{1}{r} = \frac{\theta}{1} + \frac{1-\theta}{p}$, so $\theta \in (0,1)$. By Hölder,

$$\|J_{\mu_n} x\|_{L_p(\mu_n)} \leq \|J_{\mu_n} x\|_{L_1(\mu_n)}^\theta \|J_{\mu_n} x\|_{L_p(\mu_n)}^{1-\theta}$$

Then

$$\begin{aligned}
\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_n}x\|_{L_p(\mu_n)} &\leq C\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_{n+1}}x\|_{L_r(\mu_{n+1})}\\
&\leq C\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_{n+1}}x\|_{L_1(\mu_{n+1})}^{\theta}\|J_{\mu_{n+1}}x\|_{L_p(\mu_{n+1})}^{1-\theta}\\
&\leq C\Bigg(\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_{n+1}}x\|_{L_1(\mu_{n+1})}\Bigg)^{\theta}\Bigg(\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_{n+1}}x\|_{L_p(\mu_{n+1})}\Bigg)^{1-\theta}\\
&\leq C\left(2\|J_\lambda x\|_{L_1(\lambda)}\right)^{\theta}\Bigg(\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_{n+1}}x\|_{L_p(\mu_{n+1})}\Bigg)^{1-\theta}
\end{aligned}$$

This implies that

$$\left(\frac{1}{2}\|J_{\mu_0}x\|_{L_p(\mu_0)}\right)^{\theta}\leq\Bigg(\sum_{n\geq 0} 2^{-n-1}\|J_{\mu_{n+1}}x\|_{L_p(\mu_{n+1})}\Bigg)^{\theta}\leq 2c(\|J_\lambda x\|_{L_1(\lambda)})^{\theta}$$

($c=C^{\theta}$) and

$$\|J_{\mu_0}x\|_{L_p(\mu_0)}\leq 2(2c)^{1/\theta}\,\|J_\lambda x\|_{L_1(\lambda)}$$

as desired.

□

This proof is particularly tricky, having to iterate applications of Pietch domination. It would be interesting if there were a more direct proof that could give more insight into the structure of the problem.

Week 8 (10/26/2010)

Grothendieck's Inequality

Theorem 54. (Grothendieck) *Every bounded operator from $l_1\to l_2$ is 1-summing.*

Proof. First we gather some helpful facts. Let $\{g_i\}_{i=1}^{\infty}$ be i.i.d standard Gaussian defined on a probability space (Ω,μ), Define G_p: $l_2\to L_p(\mu)$ by

$$G_p((a_i)_{i=1}^{\infty})=\sum_{i=1}^{\infty} a_i\,g_i\overset{(d)}{\sim}\|a\|_{l_2}\,g,\ g\text{ standard Gaussian}$$

For $a\in l_2$, we have

$$\|G_p\,a\|_p=\left(\mathbb{E}|\|a\|_2\,g|^p\right)^{1/p}=C_p\,\|a\|_2$$

(we computed $C_p=(\mathbb{E}|g|^p)^{1/p}$ in Lemma 31). Examining G_1: $l_2\to L_1(\Omega)$ we note that G_1 is one to one from the equality above ($\|G_pa\|_p=0$ implies $\|a\|_2=0$), and this implies that the adjoint G_1^*: $L_\infty(\Omega)\to l_2$ is onto.

Reminder: For any $a\in l_2^*$, can find $y^*\in(L_1^*)$ with $G_1^*(y^*)(x)=y^*(G_1(x))=a(x)$ for all $x\in l_2$. G_1 being one to one means that we can define such a y^* on $G_1(l_2)$, and then we extend with Hahn Banach while preserving the norm.

Then by the open mapping theorem,

$$G_1^*(B_{L_\infty(\Omega)}) \supset c B_{l_2}$$

for some c, and in fact we can show that $c \leq C_1 = \sqrt{\frac{2}{\pi}}$. Suppose the inclusion does not hold, then there exists $x \in l_2$ with $\|x\|_2 \leq c$ and $x \notin G_1^*(B_{L_\infty(\Omega)})$. By separation theorem, there exists $y \in l_2 \backslash \{0\}$ such that

$$\langle x, y\rangle > \sup_{\|f\|_\infty \leq 1} \langle y, G^*(f)\rangle$$

Then

$$c\|y\|_2 \geq \|x\|_2 \|y\|_2 \geq \langle x, y\rangle > \sup_{\|f\|_\infty \leq 1} \langle y, G^*(f)\rangle = \|G y\|_1 = C_1 \|y\|_2$$

So if the inclusion does not hold, then $c > C_1$, and thus if $c \leq C_1$, then the inclusion holds, and we have shown

Fact 1. $G_1^*(B_{L^\infty(\Omega)}) \supset c B_{l_2}$ with $c = C_1 = \sqrt{\frac{2}{\pi}}$

Now let e_1, e_2, ... be the standard basis of l_1. Noting that $c\frac{T e_i}{\|T e_i\|_2} \in c B_{l_2} \subset G_1^*(B_{L^\infty(\Omega)})$, from Fact 1 we know that there exists $x_i \in L^\infty(\mu)$ such that $G_1^*(x_i) = T e_i$ and $\|x_i\|_\infty \leq \frac{1}{c} \|T e_i\|_2$.

Define $S: l_1 \to l_\infty(\Omega)$ by $S e_i = x_i$, then for $a = (a_i)_{i=1}^\infty \in l_1$ we have

$$\|S a\|_\infty = \left\| \sum_{i=1}^\infty a_i S e_i \right\|_{L^\infty} \leq \sum_{i=1}^\infty |a_i| \|x_i\|_\infty \leq \frac{1}{c} \|T\| \|a\|_{l_1}$$

so that $\|S\| \leq \frac{\|T\|}{c}$, and we have shown

Fact 2. There exists $S: l_1 \to L_\infty(\Omega)$ such that $\|S\| \leq \frac{\|T\|}{c}$ and $T = G_1^* S$.

Recall that the identity map J: $L_\infty(\Omega) \to L_r(\Omega)$ satisfies $\pi_r(J) \leq 1$ (Proposition 48). Observe that $G_1^* = G_{r^*}^* J$ since $G_1 = J^* G_{r^*}$, and then by the ideal property

$$\pi_r(G_1^*) \leq \|G_{r^*}^*\| \pi_r(J) \leq C_{r^*}$$

In any case, we have shown the final fact that we need:

Fact 3. For $1 < r \leq 2$, we have $\pi_r(G_1^*) < \infty$.

Combining the facts, we have shown that

$$\pi_r(T) = \pi_r(G_1^* S) \leq \pi_r(G_1^* S) \leq \frac{\pi_r(G_1^*) \|T\|}{c} < \infty$$

for $1 < r \leq 2$. Then

$$\Pi_r(l_1, l_2) \subseteq \Pi_2(l_1, l_2) \subseteq \mathcal{L}(l_1, l_2) \subseteq \Pi_r(l_1 \, l_2)$$

using the inclusion of p-summing spaces Lemma 45, the fact that all p-summing operators are bounded $\|T\| \leq \pi_p(T)$, and what we have just shown: $\pi_r(T) < \infty$. Thus we have that $\Pi_r(l_1, l_2) = \Pi_2(l_1, l_2) = \mathcal{L}(l_1, l_2)$ and by Maurey Extrapolation Theorem (Theorem 53) we conclude that

$$\Pi_1(l_1, l_2) = \Pi_2(l_1, l_2) = \mathcal{L}(l_1, l_2)$$

$\square$

Theorem 55. (Extension property of 2-summing) *Let X, Y, Z be Banach spaces, $X \subseteq Z$, $\dim X < \infty$ and $T: X \to Y$ is 2-summing. Then for all $\varepsilon > 0$, there exists an extension $\tilde{T}: Z \to Y$ with*

$$\pi_2(\tilde{T}) \leq (1+\varepsilon)\pi_2(T)$$

Remark: The statement holds without the restriction that dim $X < \infty$ and without ε, but the proof needs more work.

Proof. Let $\mathcal{N}$ be an ε-net in the sphere of X^*, $|\mathcal{N}| < \infty$. Define a norm $|\cdot|_{\mathcal{N}}$ on X by

$$|x|_{\mathcal{N}} = \max_{x^* \in \mathcal{N}} |x^*(x)|$$

Also, we note that

$$\|x\| = \sup_{\|x^*\|=1} |x^*(x)| \leq \sup_{y^* \in \mathcal{N}} |y^*(x)| + \varepsilon \|x\| = |x|_{\mathcal{N}} + \varepsilon \|x\|$$

and

$$\|x\| = \sup_{\|x^*\|=1} |x^*(x)| \geq \sup_{y^* \in \mathcal{N}} |y^*(x)| = |x|_{\mathcal{N}}$$

so that $|x|_{\mathcal{N}} \leq \|x\| \leq \frac{1}{1-\varepsilon}|x|_{\mathcal{N}}$. By the ideal property, it is enough to prove the theorem when $\|x\| = |x|_{\mathcal{N}}$ (Id: $(X, \|\cdot\|) \to (X, |\cdot|_{\mathcal{N}})$ will give $1+\varepsilon$ factor). In other words, $\mathcal{N}$ is a norming set of X. By Pietch we have a probability measure μ on $\mathcal{N}$ so that the following diagram commutes. Note that X is finite dimensional so that $J(X) = C(\mathcal{N}) = l_\infty(\mathcal{N})$, and in fact $Jx = (x^*(x))_{x^* \in \mathcal{N}}$.

$$\begin{array}{ccccc} & & l_\infty(\mathcal{N}) & \xrightarrow[I]{} & L_2(\mu) \\ & \tilde{J} \nearrow & J\uparrow & & \downarrow S \\ Z & \supseteq & X & \xrightarrow{T} & Y \end{array}$$

Here $\|S\| \leq \pi_2(T)$ and $\pi_2(I) \leq 1$. By Hahn-Banach, there exists $\tilde{x}^* \in Z^*$ such that $\tilde{x}^*|_X = x^*$ and $\|\tilde{x}^*\| = \|x^*\|$. Define $\tilde{J}(z) = (\tilde{x}^*(z))_{x^* \in \mathcal{N}}$, and by construction $\|\tilde{J}\| = \|J\| = 1$. Now taking $\tilde{T} = SI\tilde{J}$, we note that $\tilde{T}|_X = T$ by construction, and ideal property tells us that $\pi_2(\tilde{T}) \leq \pi_2(T)$, which gives the result (recall we lost a factor of $1+\varepsilon$ in replacing $(X, \|\cdot\|)$ by $(X, |\cdot|_{\mathcal{N}})$).

$\square$

We can use the previous result to prove a theorem about projections to finite dimensional subspaces.

Theorem 56. (Kadets-Snobar) *$X \subseteq Z$ Banach spaces, $\dim X = n$. For all $\varepsilon > 0$, there exists a projection $P: Z \to X$ with*

$$\|P\| \leq \sqrt{n}(1+\varepsilon)$$

Remark: $\|P\| \leq \sqrt{n}$ works also. This is not sharp either, as a result by Konig-Tomczak-Jaegerman can find $\|P\| \leq \sqrt{n} - \frac{c}{\sqrt{n}}$.

Proof. We apply the previous 2-sum extension result Theorem 55 to the identity map $I_X: X \to X$, which gives some operator $P: Z \to X$ with $P|_X = I_X$ and $\pi_2(P) \leq (1+\varepsilon)\pi_2(I_X)$. We know that $\pi_2(I_X) = \sqrt{n}$ (Theorem 51). We conclude that

$$\|P\| \leq \pi_2(P) \leq (1+\varepsilon)\sqrt{n}$$

$\square$

(Later, after we finish p-summing results, combined with Dvoretsky we can show a theorem by Lindenstrauss-Tzafriri: If Z is a Banach space such that for all closed $X \subseteq Z$ there exists a projection $P: Z \to X$ with $\|P\| < \infty$ then Z is essentially a Hilbert space: there exists a scalar product $\langle \cdot, \cdot \rangle$ on Z such that $\sqrt{\langle x, x\rangle} \leq \|x\| \leq K\sqrt{\langle x, x\rangle}$)

Theorem 57. (Grothendieck's Inequality) *There exists K such that for any $n \times n$ real matrix (a_{ij}),*

$$\max_{\substack{x_1\ldots,x_n,\\ y_1,\ldots,y_n \in l_2\\ \text{unit vectors}}} \sum_{i,j=1}^n a_{ij}\langle x_i, x_j\rangle \leq K \max_{\substack{\varepsilon_1,\ldots,\varepsilon_n,\\ \delta_1,\ldots,\delta_n \in \{\pm 1\}}} \sum_{i,j=1}^n a_{ij}\,\varepsilon_i\,\delta_j$$

The best K in this inequality is denoted K_G, Grothendieck's constant.

Remark: Note that

$$\max_{\varepsilon_i,\delta_j=\pm 1} \sum_{i,j=1}^n a_{ij}\,\varepsilon_i\,\delta_j = \max_{\substack{(\alpha,\beta)\in\mathbb{R}^n\times\mathbb{R}^n\\ |\alpha_i|,|\beta_j|\leq 1}} \sum_{i,j=1}^n a_{ij}\,\alpha_i\,\beta_j$$

Can optimize coordinate wise to see that the maximizer is achieved at the endpoints (LHS).

Proof. Denote

$$\begin{aligned}
\Gamma &= \max_{\substack{\|x_i\|_2\leq 1\\ \|y_j\|_2\leq 1}} \sum_{i,j=1}^n a_{ij}\langle x_i, y_j\rangle \\
\Delta &= \max_{\varepsilon_i,\delta_j=\pm 1} \sum_{i,j=1}^n a_{ij}\,\varepsilon_i\,\delta_j
\end{aligned}$$

Goal is to show that $\Gamma = O(\Delta)$. Let $\{g_i\}_{i=1}^{\infty}$ be i.i.d standard Gaussians defined on the same probaiblity space (Ω, μ).

$$H := G_2(l_1) := \left\{ \sum_{i=1}^{\infty} a_i\,g_i : (a_i) \in l_2 \right\}$$

H is a Hilbert space with the scalar product $\langle X, Y\rangle = \mathbb{E}[XY]$.

Fix $\varepsilon > 0$. Let $X_1 \ldots, X_n, Y_1, \ldots, Y_n \in H$ be such that $\mathbb{E}[X_i^2], \mathbb{E}[Y_i^2] \leq 1$ and

$$\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[X_i Y_j] \geq \Gamma - \varepsilon$$

We will be moving between $l_2, H, L_2(\mu)$, equivalent Hilbert spaces.

For $M>0$ to be determined, and any $X\in L_2(\mu)$, define

$$X^M=\begin{cases} X & \text{if } |X|\le M\\ M & \text{if } X>M\\ -M & \text{if } X<-M\end{cases}$$

so that $|X^M|\le M$. then

$$\Gamma-\varepsilon\le\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[X_iY_j]=\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[X_i^MY_j^M]+\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[(X_i-X_i^M)Y_j]+\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[X_i^M(Y_j-Y_j^M)]$$

Bounding each piece on the RHS, we have that

$$\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[X_i^MY_j^M]=M^2\mathbb{E}\left[\sum_{i,j=1}^n a_{ij}\frac{X_i^M}{M}\frac{Y_j^M}{M}\right]\le M^2\max_{\|\alpha\|_\infty,\|\beta\|_\infty\le 1}\sum_{i,j=1}^n a_{ij}\alpha_i\beta_j\le M^2\,\Delta$$

(using the remark). For the second part,

$$\begin{aligned}\sum_{i,j=1}^n a_{ij}\,\mathbb{E}[(X_i-X_i^M)Y_j] &= \max_i\|X_i-X_i^M\|_{L_2(\mu)}\sum_{i,j=1}^n a_{ij}\,\mathbb{E}\left[\frac{(X_i-X_i^M)}{\max_i\|X_i-X_i^M\|_{L_2(\mu)}}Y_j\right]\\ &\le \max_i\|X_i-X_i^M\|_{L_2(\mu)}\Gamma\end{aligned}$$

where for the inequality we are using the equivalence of H and l_2 as Hilbert spaces. Now since $X_i=\sum a_jg_j$, $\sum a_j^2=\mathbb{E}[X_i^2]=1$ so $X_i\sim\|a\|_2g\sim N(0,1)$, and therefore

$$\begin{aligned}\sqrt{\mathbb{E}[(X_i-X_i^M)^2]} &= \sqrt{\int_0^\infty t\,\Pr(|X_i-X_i^M|>t)\,dt}\\ &= \sqrt{2\int_M^\infty (t-M)\Pr(X_i>t)\,dt}\\ &\le Ce^{-M^2/4}\end{aligned}$$

using the standard Gaussian tail estimate. Thus both the second and third terms above are bounded by $Ce^{-M^2/4}\Gamma$, and

$$\Gamma-\varepsilon\le M^2\,\Delta+Ce^{-M^2/4}\Gamma$$

so that

$$\Gamma(1-Ce^{-M^2/4})\le M^2\Delta$$

and now we just choose M sufficiently large so that $\Gamma\le C'(M)\Delta$.

□

Theorem 58. (Grothendieck) *Any bounded operator $T: l_\infty\to l_1$ is 2-summing and*

$$\pi_2(T)\le K_G\,\|T\|$$

where K_G is Grothendieck's constant.

Also any bounded operator from $l_1\to l_2$ satisfies $\pi_1(T)\le K_G\,\|T\|$.

Remark: Can replace l_∞ and l_1 with $L_\infty(\mu)$ and $L_1(\mu)$ as in the statement of Maurey Extrapolation (Theorem 53)

Proof. It is enough to prove this for $T\colon l_\infty^n \to l_1^n$, as the 2-summing property involves finite sums. Write $Te_i = \sum_{j=1}^n a_{ij} e_j$, and take $x_1, \ldots, x_m \in l_\infty^n$, $x_k = \sum_{i=1}^n \alpha_i^k e_i$ such that

$$\sup_{\|x^*\|_{l_1} \leq 1} \sqrt{\sum_{i=1}^n |x^*(x_i)|^2} = 1$$

We need to show that $\sqrt{\sum_{i=1}^m \|Tx_k\|^2} \leq K_G \|T\|$. Note that

$$\max_{i=1,\ldots,n} \sum_{k=1}^n (\alpha_i^k)^2 \leq \sup_{\|x^*\|_{l_1} \leq 1} \sqrt{\sum_{i=1}^n |x^*(x_i)|^2} = 1$$

taking x^* to be standard basis elements. Now

$$\begin{aligned} \sqrt{\sum_{i=1}^m \|Tx_k\|_1^2} &= \left(\sum_{k=1}^m \left(\sum_{j=1}^n \left| \sum_{i=1}^n \alpha_i^k \, a_{ij} \right| \right)^2 \right)^{1/2} \\ \text{(Minkowski)} &\leq \sum_{j=1}^n \left(\sum_{k=1}^m \left| \sum_{i=1}^n \alpha_i^k \, a_{ij} \right|^2 \right)^{1/2} =: A \end{aligned}$$

By duality, there exists $y_1, \ldots, y_n \in l_2^m$, $y_j = (\beta_j^k)_{k=1}^m$, with $\|y_j\|_{l_2} \leq 1$ and

$$\begin{aligned} A &= \sum_{j=1}^n \sum_{k=1}^m \beta_j^k \sum_{i=1}^n \alpha_i^k a_{ij} \\ &= \sum_{i,j=1}^n \left(\sum_{k=1}^m \alpha_i^k \beta_j^k \right) a_{ij} \\ &= \sum_{i,j=1}^n \langle \alpha, \beta \rangle \, a_{ij} \\ \text{(Grothendieck)} &\leq K_G \max_{|\alpha_i| \leq 1, |\beta_j| \leq 1} \sum_{i,j=1}^n a_{ij} \alpha_i \beta_j \\ &= K_G \max_{\|\alpha\|_\infty, \|\beta\|_\infty \leq 1} \langle T\alpha, \beta \rangle \\ &= K_G \, \|T\|_{\infty \to 1} \end{aligned}$$

which is the result we want (slight abuse of notation above, at the application of Grothendieck we have reused the variable α, β).

(Next time to prove for $T\colon l_1 \to l_2$) □

Week 9 (11/2/2010)

We now prove the next part, that any bounded operator from $l_1 \to l_2$ is 1-summing.

Proof. As before, it is enough to prove this for $T\colon l_1^n \to l_2^n$. We need that for any $x_1, \ldots, x_m \in l_1$,

$$\sum_{i=1}^m \|Tx_i\|_2 \leq K_G \|T\| \sup_{|\alpha_i| \leq 1} \sum |\alpha(x_i)|$$

We will use that $\sup_{|\alpha_i|\le 1}\sum|\alpha(x_i)| = \sup_{|\alpha_i|\le 1}\|\sum\alpha_i x_i\|_{l_1^n}$ (Proposition 46). Let $x_i = (x_{i1}, ..., x_{in})$, and $e_1..., e_n$ be the standard basis of l_1^n. Then

$$\begin{aligned}
\sum_{i=1}^{\infty}\|Tx_i\|_2 &= \sum_{i=1}^{m}\left\|\sum_{j=1}^{n} x_{ij}Te_j\right\| \\
&= \sum_{i=1}^{m}\left\langle z_i, \sum_{j=1}^{n} x_{ij}Te_j\right\rangle \text{ for some } \|z_i\| = 1 \\
&= \|T\|\sum_{i=1}^{m}\sum_{j=1}^{n}\left\langle z_i, \frac{Te_j}{\|T\|}\right\rangle x_{ij} \\
\text{(Grothendieck)} \quad &\le \|T\|K_G \sup_{|\alpha_i|, |b_j|\le 1}\sum_i\sum_j \alpha_i\beta_j x_{ij} \\
&= \|T\|K_G \sup_{|\alpha_i|, |b_j|\le 1}\sum_j \beta_j\sum_i \alpha_i x_{ij} \\
&= \|T\|K_G \sup_{|\alpha_i|\le 1}\left\|\sum_i \alpha(x_i)\right\|
\end{aligned}$$

which completes the proof. □

We can also transfer the result from $L_\infty(\mu)$ to $C(K)$ where K is compact Hausdorff, but it requires a bit of thought. For the next results K will denote a compact Hausdorff space.

Theorem 59. *If $E \subseteq C(K)$ is a finite dimensional subspace then there exists another finite dimensional subspace $F \subseteq C(K)$ such that*

1. *$E \subseteq F$*
2. *F is $(1+\varepsilon)$-isomorphic to $l_\infty^{\dim F}$*

We then say that $C(K)$ is a $\mathcal{L}_{\infty,1+\varepsilon}$ space. (Notation in literature)

This will use the following Lemma:

Lemma 60. *Let $E \subseteq C(K)$ be a finite dimensional subspace, $\varepsilon > 0$. Then there exists a projection P: $C(K) \to C(K)$ such that*

1. *P has finite rank*
2. *$\|Px - x\|_\infty \le \varepsilon$ for all $x \in B_E$*
3. *$P(X)$ is isometric to l_∞^n, where $n = \dim P(X)$.*

Proof. This is a partition of unity argument. Let $f_1, ..., f_n$ be an $\varepsilon/4$ net in B_E. Note that f_i are uniformly continuous. There exists a finite cover of K by open sets $\Omega_1, ..., \Omega_N$ such that

$$|f_i(x) - f_i(y)| < \frac{\varepsilon}{2} \text{ for all } x, y \in \Omega_j \text{ and for all } i$$

This covers K into pieces where on each piece all the f_i oscillate by at most ε. Without loss of generality, $\Omega_i \setminus \bigcup_{j\ne i}\Omega_j \ne 0$, otherwise we can toss Ω_i out of the cover. Fix a ω_i from each $\Omega_i \setminus \bigcup_{j\ne i}\Omega_j$. Let $\{\varphi_j\}_{j=1}^N$ be a partition of unity subordinate to the open cover $\{\Omega_i\}$. This means that

1. $\varphi_j \in C(K)$

2. $0 \le \varphi_j \le 1$

3. φ_j vanishes outside Ω_i

4. $\sum \varphi_i(\omega) = 1$ for all $\omega \in K$

Define $P: C(K) \to C(K)$ by

$$Pf(\omega) = \sum_i f(\omega_i)\,\varphi_i(\omega)$$

Note this is a projection, $\omega_i \in \Omega_i \setminus \bigcup_{j \neq i} \Omega_j$ means that $\varphi_i(\omega_j) = \delta_{ij}$ and $Pf(w_i) = f(\omega_i)$, so that $P^2 = P$. Also, we have that

$$|Pf(\omega)| \le \sum_i |f(\omega_i)|\,\varphi_i(\omega) \le \|f\|_\infty \sum_i \varphi_i(\omega) = \|f\|_\infty$$

so that $\|P\| \le 1$. We see that P has finite rank, since $P(C(K)) = \operatorname{span}\{\varphi_1, \ldots, \varphi_N\}$, and for all $\omega \in K$ we have that

$$\begin{aligned}
|Pf(\omega) - f(\omega)| &= \left| \sum_j (f(\omega_j) - f(\omega))\varphi_j(\omega) \right| \\
&\le \sum_j \left(\underbrace{|f(\omega_j) - f_i(\omega_j)|}_{\le \varepsilon/4} + |f_i(\omega_j) - f_i(\omega)| + \underbrace{|f(\omega) - f_i(\omega_j)|}_{\le \varepsilon/4} \right) \varphi_j(\omega) \\
&\le \frac{\varepsilon}{2} + \sum_{j, \omega \in \Omega_j} \underbrace{|f_i(\omega_j) - f_i(\omega)|}_{\le \varepsilon/2} |\varphi_j(\omega)| \\
&\le \varepsilon
\end{aligned}$$

so that $\|Pf - f\| \le \varepsilon$ for $f \in B_E$. In the second line above, note that $\sum \varphi_j = 1$ and f_i is an $\varepsilon/4$ net. In the third line above, we use that $\varphi_j(\omega)$ is zero outside Ω_j and f_i does not oscillate by more than $\varepsilon/2$ on Ω_j. Finally, to show that $P(C(K))$ is isomorphic to l_∞^N, we just map $(a_i)_{i=1}^N$ to $\sum a_i \varphi_i$, and since φ_i is a partition of unity the sup norm is exactly the same:

$$\left| \sum a_i \varphi_i \right| \le \|a\|_\infty = \left| \sum_i a_i\, \varphi(w_{i_{\max}}) \right| = \left\| \sum a_i \varphi_i \right\|_\infty$$

□

Now we prove the Theorem:

Proof. (of Theorem 59) Denote $X = C(K)$. We are given $E \subseteq X$, $\dim E = k$. By Kadetz-Snobar (Theorem 56) there exists a projection $Q: C(K) \to E$ with $\|Q\| \le 2\sqrt{k}$. By the previous lemma, there exists a projection $P: X \to X$ with $\|Px - x\| \le \varepsilon$ for all $x \in B_E$ and $P(X)$ is isometrically isomorphic to l_∞^n. Let I: $X \to X$ be the identity map, and define $T = I + PQ - Q$. Then

$$\begin{aligned}
\|I - T\| &\le \|(P - I)Q\| \\
&\le \sup_{f \in B_X\, \omega \in K} \left| P\frac{Qf(\omega)}{\|Q\|} - \frac{Qf(\omega)}{\|Q\|} \right| \cdot \|Q\| \\
&\le \varepsilon 2\sqrt{k}
\end{aligned}$$

so if $\varepsilon < \frac{1}{2\sqrt{k}}$, then T is invertible, $T^{-1} = \sum_{j \ge 0} (I - T)^j$ (Neumann series), $\|T^{-1}\| \le \sum_{j \ge 0} (2\varepsilon\sqrt{k})^j = \frac{1}{1 - 2\varepsilon\sqrt{k}}$. Also, we note $\|T\| \le 1 + \varepsilon 2\sqrt{k}$.

We use $T^{-1}PT$, which is a finite rank projection from $X \to X$. Denote the range by F. Then $F \supseteq E$ since if $x \in E$, then

$$Tx = x + PQx - Qx = Px$$

so that $x = T^{-1}PTx$ and $x \in F$. Let T_0 be the restriction of T to F. Then note that

$$T_0(F) = T_0(T^{-1}PTX) = P(TX) = PX$$

(T is invertible so $TX = X$, and $T_0 = T$ on the range of PTX), and so we have that $T_0: F \to PX$, and PX is isomorphic to l_∞^n. Note

$$\|T_0\|\,\|T_0^{-1}\| \leq \frac{1 + 2\varepsilon\sqrt{k}}{1 - 2\varepsilon\sqrt{k}} \sim 1 + \varepsilon'$$

ε' is arbitrarily small. Thus F is $1 + \varepsilon'$ isomorphic to l_∞^n.

□

From this we can then prove the following in the same manner as in remarks following Theorem 53 (Maurey Extrapolation)

Proposition 61. *Let K be compact Hausdorff, then any bounded operator $T: C(K) \to L_1(\mu)$ is two summing with $\pi_2(T) \leq K_G \|T\|$*

Proof. Let $x_1, \ldots, x_n \in C(K)$. Take $E = \operatorname{span}\{x_1, \ldots, x_n\}$. There exists $F \supseteq E$, $F \subset C(K)$ which is $1 + \varepsilon$ isomorphic to l_∞^N, i.e. $U: F \to l_\infty^N$ is invertible with $\|U\|, \|U^{-1}\| \leq 1 + \varepsilon$. If we consider the map $\tilde{T} = TU^{-1}$ mapping $l_\infty^N \to L_1$, we know from Grothendieck (Theorem 58) that $\pi_2(\tilde{T}) \leq K_G \|\tilde{T}\|$, so that

$$\sum_i \|Tx_i\|_1 = \sum_i \|\tilde{T}(Ux_i)\|_1 \leq K_G \|\tilde{T}\| \sup_{\|x^*\|_{l_1} \leq 1} \sum_i |x^*(Ux_i)| \leq K_G(1 + \varepsilon)\|T\| \sup_{y^* \in B_{C(K)^*}} \sum_i |y^*(x_i)|$$

which shows $\pi_2(T) \leq K_G(1 + \varepsilon)\|T\|$, which is slightly off.

□

We can use this theorem to prove a result about Fourier multipliers. Below we will state the result using the Fourier transform of abelian groups. Can think of just the Fourier series on $\mathbb{S}^1$ with the dual $\mathbb{Z}$ (Fourier coefficients index)

Theorem 62. (Orlicz-Paley-Sidon) *Let G be a compact abelian group, and Γ be the dual. Let M be a Fourier multiplier $M: C(G) \to \mathbb{C}^\Gamma$ with*

$$Mf = (m(\chi)\,\hat{f}(\chi))_{\chi \in \Gamma}$$

for some choice $\{m(\chi)\}_{\chi \in \Gamma} \subset \mathbb{C}$. Then $M(C(G)) \subseteq l_1(\Gamma)$ if and only if $\sum_{\chi \in \Gamma} |m(\chi)|^2 < \infty$.

Proof. First assume that $\sum_{\chi \in \Gamma} |m(\chi)|^2 < \infty$. Then

$$\sum_{\chi \in \Gamma} |m(\chi)|\,|\hat{f}(\chi)| \leq \sqrt{\sum_{\chi \in \Gamma} |m(\chi)|^2}\sqrt{\sum_{\chi \in \Gamma} |\hat{f}(\chi)|^2} = \sqrt{\sum_{\chi \in \Gamma} |m(\chi)|^2}\, \|f\|_{L_2(\mu)} < \infty$$

where μ is the Haar measure on G. This shows that $M(C(G)) \subseteq l_1(\Gamma)$.

The converse is true, and the proof uses Grothendieck's results. Assume that $M(C(G)) \subseteq l_1(\Gamma)$. Check that (f, Mf) is a closed graph, which will then imply that $\|M\|_{C(G) \to l_1(\Gamma)} < \infty$. By Grothendieck (Theorem 58), this implies that M is 2-summing with $\pi_2(M) \leq K_G \|M\|$. By Pietch domination, this implies that there exists a Borel probability measure ν on G such that

$$\sum_{\chi \in \Gamma} |m(\chi)| |\hat{f}(\chi)| \leq K_G \|M\| \left(\int_G |f(\chi)|^2 \, d\nu(\chi) \right)^{1/2}$$

Let $h \in G$ and $f \in C(G)$. Define $f_h(g) = f(hg)$. We have that

$$\begin{aligned} \hat{f_h}(\chi) &= \int_G f(hx) \, \chi(x) \, d\mu(x) \\ &= \int_G f(y) \, \chi(h^{-1} y) \, d\mu(y) \\ &= \chi(h^{-1}) \, \hat{f}(y) \end{aligned}$$

Above we used the invariance of the Haar measure under group multiplication and the fact that χ is a character (homomorphism from $G \to \mathbb{C}$). Now $|\hat{f_h}(\chi)| = |\hat{f}(\chi)|$ for all $\chi \in \Gamma$, and

$$\begin{aligned} \left(\sum_{\chi \in \Gamma} |m(\chi)| |\hat{f}(\chi)| \right)^2 &\leq K_G^2 \|M\|^2 \int_G \int_G |f(hx)|^2 \, d\nu(x) \, d\mu(h) \\ &= K_G^2 \, \|M\|^2 \int_G \left(\int_G |f(hx)|^2 \, d\mu(h) \right) d\nu(x) \\ &= K_G^2 \|M\|^2 \, \|f\|_{L_2(\mu)}^2 \\ \text{(Parseval)} \quad &= K_G^2 \, \|M\|^2 \sum_{\chi \in \Gamma} |\hat{f}(\chi)|^2 \end{aligned}$$

For all finite $\Gamma_0 \subseteq \Gamma$, we have $f_0 = \sum_{\chi \in \Gamma_0} m(\chi) \cdot \chi \in C(K)$. Then

$$\begin{aligned} \left(\sum_{\chi \in \Gamma_0} |m(\chi)|^2 \right)^2 &\leq K_G^2 \, \|M\|^2 \sum_{\chi \in \Gamma_0} |m(\chi)|^2 \\ \sum_{\chi \in \Gamma_0} |m(\chi)|^2 &\leq K_G^2 \|M\|^2 < \infty \end{aligned}$$

using the fact that $\hat{\chi} = \delta_\chi$ so that $\hat{f}_0(\chi) = m(\chi)$.

$\square$

We finish this section with another theorem of Grothendieck.

Theorem 63. (Grothendieck) *Let X be a closed linear subspace of L_1. Assume that X^* is K-isomorphic to a subspace of L_1. Then X is $K_G K$-isomorphic to a Hilbert space.*

An example of a subspace of L_1 which is isomorphic to a Hilbert space is the space of random variables $\left\{ \sum a_i g_i, \ \sum |a_i|^2 < \infty \right\}$ where g_i are i.i.d $N(0,1)$ gaussian and $\|\sum a_i g_i\|_{L^1} = \sqrt{\sum |a_i|^2} \, E|g|$. Note it is a Hilbert space with inner product $\langle f, g \rangle = \mathbb{E}[fg]$. (We used this in the proof of Grothendieck's inequality, Theorem 57)

Proof. For the convenience of proof, we will use $L_1 = L_1([0,1])$. Denote $I\colon X \to L_1$ by the identity, noting the adjoint $I^*\colon L_\infty \to X^*$. There exists $T\colon X^* \to L^1$ with $\|T\| \le 1$, $\|T^{-1}\| \le K$ by assumption. Then TI^*: $L_\infty \to L_1$, so using Theorem 58 (Bounded operators from $L_\infty \to L_1$ are 2-summing),

$$\pi_2(TI^*) \le K_G \, \|TI^*\| \le K_G$$

By Pietch Domination, we can express TI^* as a composition of maps through a Hilbert space H:

$$\begin{tikzcd} & H \arrow[dr, "A"] & \\ L_\infty \arrow[ur, "B"] \arrow[r, "I^*"] & X^* \arrow[r, "T"] & L_1 \end{tikzcd}$$

with $\|B\| \le 1$ and $\|A\| \le \pi_2(TI^*) \le K_G$. We know that $TX^* \subset L_1$. Denote $A^{-1}(TX^*)$ by $H_0 \subset H$. Let P: $H \to H_0$ be the orthogonal projection. The diagram becomes

$$\begin{tikzcd} & H \arrow[r, "P"] & H_0 \arrow[d, "A"] & \\ L_\infty \arrow[ur, "B"] \arrow[r, "I^*"] & X^* \arrow[r, "T"] & TX^* & \subseteq L_1 \end{tikzcd}$$

and defining $\tilde{B} = PB\colon L_\infty \to H_0$ and $\tilde{A} = T^{-1}A\colon H_0 \to X^*$ we have

$$\begin{tikzcd} & H_0 \arrow[dr, "\tilde{A}"] & \\ L_\infty \arrow[ur, "\tilde{B}"] \arrow[rr, "I^*"] & & X^* \end{tikzcd}$$

with $\|\tilde{B}\| \le 1$ and $\|\tilde{A}\| \le K_G K$. Now we "dualize"...

$$\begin{tikzcd} & H_0 \arrow[dl, "\tilde{B}^*"'] & \\ L_\infty^* & & X^{**} \arrow[ll, "I^{**}"'] \arrow[ul, "\tilde{A}^*"'] \end{tikzcd}$$

Note $X \subset X^{**}$ and $I^{**}|_X = I$. We restrict everything to X:

$$\begin{tikzcd} & H_0 \arrow[dl, "V"'] & \\ L_1 & & X \arrow[ll, "I"'] \arrow[ul, "U"'] \end{tikzcd}$$

where $U = \tilde{A}^*\Big|_X$ and $V = \tilde{B}^*\Big|_{UX}$. Restrictions of operators have the same operator norm, and we have that

$$\|x\| = \|Ix\| = \|VUx\| \le \|Ux\| \le K_G K \, \|x\|$$

or

$$\|x\| \le \|Ux\| \le K_G K \, \|x\|$$

which shows that X is $K_G K$ isomorphic to $UX \subseteq H_0$, a Hilbert space. □

This kind of argument can go on for many semesters. Now we change to another topic.

Lindenstrauss-Tzafriri Theorem

The last part of the course will be leading towards a theorem of Lindenstrauss and Tzafriri: Let X be a Banach space such that any closed subspace of X has a bounded projection P onto Y (if this is the case we say that Y is **complemented**). Then X is K-isomorphic to a Hilbert space for some $K<\infty$.

This is actually a theorem about finite dimensions, though not immediately obvious. The proof involves a clever use of Dvoretsky's theorem. We need some preparation first.

Theorem 64. *Assume that all closed subspaces of X are complemented. Then there exists $\lambda<\infty$ such that any finite subspace of X is complemented via a projection of norm $\leq\lambda$.*

The proof involves the following Lemma:

Lemma 65. *Let $\varepsilon>0$, X Banach space, $E\subseteq X$ finite dimensional subspace. Then there exists a finite codimensional subspace $X_0\subseteq X$ such that $\|e+x\|\geq(1-\varepsilon)\|e\|$ for $e\in E, x\in X_0$.*

The conclusion states that we can find a projection from $E+X_0\to E$ with norm bounded by $\frac{1}{1-\varepsilon}$. By replacing X_0 with some closed complement of $X_0\cap E$ in X_0 (for any finite dimensional space there exists a closed complement), we can arrange it so that $E\cap X_0=\{0\}$ so that $E+X_0=E\oplus X_0$.

Proof. Let $x_1,\ldots,x_n$ be an ε-net in the unit sphere of E. For all i, let $x_i^*\in X^*$ be such that $\|x_i^*\|\leq1$ and $x_i^*(x_i)=1$ (Hahn-Banach). Then we use $X_0=\bigcap_{i=1}^n \ker(x_i^*)$. X_0 has finite codimension, being the finite intersection of codimension-1 subspaces. For $\|e\|=1$, $e\in E$, there exists i such that $\|x_i-e\|\leq\varepsilon$. Take $x\in X_0$, so that $x_i^*(x)=0$. Then

$$\|e+x\|\geq\|x_i+x\|-\|e-x_i\|\geq x_i^*(x_i+x)-\varepsilon=1-\varepsilon$$

□

Notation: For finite dimensional subspaces $E\subseteq X$ let $\lambda(E)$ be the smallest norm of a projection $P\colon X\to E$, which exists by a compactness argument (exercise: set of projections from $X\to E$ form a compact set, and the norm $P\mapsto\|P\|$ is continuous).

Then the conclusion of the theorem states that

$$\sup\{\lambda(E)\colon E\subseteq X, \dim E<\infty\}<\infty$$

In the proof of the theorem we will be assuming the opposite and find a contradiction. First we show the following reduction:

Lemma 66. *Assume that*

$$\sup\{\lambda(E)\colon E\subseteq X, \dim E<\infty\}=\infty$$

Then for all finite codimensional $X_0\subseteq X$,

$$\sup\{\lambda(E)\colon E\subseteq X_0, \dim E<\infty\}=\infty$$

Proof. Suppose that for some finite codimensional X_0,

$$\sup\{\lambda(E)\colon E\subseteq X_0, \dim E<\infty\}=M<\infty$$

Let $k = \operatorname{codim} X_0$. Take an arbitrary finite dimensional subspace $E \subseteq X$. $E_0 = E \cap X_0$. Let $P_0: X \to E_0$ be a projection $\|P_0\| \leq M$. Define $F = \{x \in E: P_0 x = 0\}$, i.e. $E = E_0 \oplus F$. Then since the codimension of $X_0 \leq k$, $\dim F \leq k$. Kadets-Snobar (Theorem 56) gives some projection $P_1: X \to F$ with $\|P_1\| \leq 2\sqrt{k}$. Now look at $P = P_0 + P_1 - P_1 P_0: X \to E$. First, P is a projection on E. For $x \in E$, we can decompose $x = a + b$ with $a \in E_0$ and $b \in F$. Then

$$Px = Pa + Pb = a + b$$

and

$$\|P\| \leq M + 2\sqrt{k} + M2\sqrt{k}$$

which shows that

$$\sup \{\lambda(E): E \subseteq X, \dim E < \infty\} \leq M + 2\sqrt{k} + M2\sqrt{k}$$

contradicting the initial assumption. □

Now we have all the ingredients to prove the theorem:

Proof. (of Theorem 64)

(Adapted from Theorem 1 of Davis, Dean, Singer. *Complemented Subspaces and* Λ *Systems in Banach Spaces*.)

Assume towards a contradiction that $\sup \{\lambda(E): E \subseteq X, \dim E < \infty\} = \infty$. The goal is to construct a closed subspace of X which is not complemented.

First, choose $E_1 \subset X$ with $\lambda(E_1) \geq 1$. By Lemma 65, we can find a finite co-dimensional subspace X_1 for which the natural projection from $E_1 \oplus X_1 \to E_1$ has norm bounded by 2.

Since X_1 is finite co-dimensional, we apply Lemma 66 which tells us that $\sup \{\lambda(E): E \subseteq X_0, \dim E < \infty\} = \infty$ also, and thus we can find a finite dimensional subspace $E_2 \subset X_1$ for which $\lambda(E_2) \geq 2$. Using the same argument as for E_1 above, there exists a finite co-dimensional subspace $X_2 \subset X_1$ for which the natural projection from $E_1 \oplus E_2 \oplus X_2 \to E_1 \oplus E_2$ has norm bounded by 2.

Continuing in this way, we have a sequence of spaces E_n and X_n for which $\lambda(E_n) \geq n$, $X_{n+1} \subset X_n$ and the natural projections of $E_1 \oplus \ldots \oplus E_n \oplus X_n$ onto $E_1 \oplus \ldots \oplus E_n$ have norm bounded by 2.

Define

$$F = \sum_{n=1}^{\infty} E_n := \left\{ \sum e_n \text{ s.t. } e_n \in E_n \text{ and } \sum e_n \text{ converges in } X \right\}$$

F is a closed subspace of X (exercise). Let P_n be the natural projection of F onto $E_1 \oplus \ldots \oplus E_n$. Suppose P is a projection of X onto F. Then $P_n P - P_{n-1} P_n P$ is a projection of X onto E_n with norm bounded by $\|P_n\| \|P\| + \|P_{n-1}\| \|P_n\| \|P\| \leq 6\|P\|$, so that $\lambda(E_n) \leq 6\|P\|$. This contradicts $\lambda(E_n) \geq n$ for sufficiently large n, and thus there does not exist a bounded projection from X onto F (F is not complemented).

□

Week 10 (11/9/2010)

Notation: Let d_X denote the distance from X to a Hilbert space:

$$d_X = \inf \{K, \text{ exists } T: X \to H \text{ s.t. } \|x\| \leq \|Tx\| \leq K\|x\|\}$$

The Lindenstrauss-Tzafriri Theorem then says that if X is a Banach space such that for all closed linear subspaces are complemented, then $d_X < \infty$.

John's Theorem: If $\dim X = n < \infty$, then $d_X \leq \sqrt{n}$.

Last time we proved the first ingredient of the Lindenstrauss-Tzafriri theorem in Theorem 64. Here is another ingredient we will need:

Lemma 67. *Let $C > 0$, and X be a separable Banach space such that for all finite dimensional subspaces $E \subset X$ we have that $d_E \leq C$. Then $d_X \leq C$.*

Remark: The statement holds for non-separable as well, and the proof can be carried out the same way replacing sequences with nets.

Proof. Let $E_1 \subseteq E_2 \subseteq \ldots$ be finite dimensional subspaces such that $E = \bigcup_{i=1}^{\infty} E_i$ is dense in X. For all n, there exists a Hilbertian norm $\|\cdot\|_n$ on E_n such that for all $n \in \mathbb{N}$ and $x \in E_n$, we have

$$\|x\| \leq \|x\|_n \leq C\,\|x\|$$

There exists a subsequence n_k such that for all $x \in E$, $\{\|x\|_{n_k}\}_{k=1}^{\infty}$ converges. This is done using the usual diagonalization argument, first for a countable dense subset, and extending to the rest of E by density, $\|x\|_\infty := \lim_{k\to\infty} \|x\|_{n_k}, x \in E$. Then to show that the limiting norm is Hilbertian we show that it satisfies the parallelogram identity

$$\|x + y\|_\infty^2 + \|x - y\|_\infty^2 = 2\|x\|_\infty^2 + 2\|x\|_\infty^2$$

$\square$

Theorem 68. *Let X is an infinite dimensional Banach space. Let $E \subseteq X$ be a finite dimensional subspace. Then for all $n \in \mathbb{N}$, and for all $\varepsilon \in (0,1)$, there exists a norm $\|\cdot\|_Y$ on $E \oplus l_2^m$ such that*

1. *There exists a linear map $R: Y \to X$ such that $\|Rx\| = \|x\|_Y$ for all $x \in Y$*
2. *$\|(x,0)\|_Y = \|x\|$ for all $x \in E$*
3. *$(1-\varepsilon)\,\|z\| \leq \|(0,z)\|_Y \leq \|z\|$ for all $z \in l_2^m$.*
4. *$(1-\varepsilon)\,\|(x,-z)\|_Y \leq \|(x,z)\|_Y \leq (1+\varepsilon)\,\|(x,-z)\|_Y$*

The first three properties essentially is the statement of Dvoretsky's Theorem, finding a subspace that is $1+\varepsilon$ isomorphic to a Hilbert space. But we can do this in such a way that we get an additional symmetry.

Proof. Fix $\delta \in (0,1)$, $n > m$, so that $l_2^m \subset l_2^n$. Let $\{x_j\}_{j=1}^N$ be a δ-net in B_E, and let $\{z_j\}_{j=1}^M$ be an ε-net in $\mathbb{S}^{n-1}$. By Dvoretsky's Theorem, there exists a linear $S: l_2^n \to X$ such that for all $z \in l_2^n$,

$$(1-\delta)\,\|z\| \leq \|Sz\| \leq \|z\|$$

For $i \leq j \leq N$, $1 \leq k \leq \frac{1}{\delta}$ define $f_{j,k}: \mathbb{S}^n \to \mathbb{R}$ by

$$f_{j,k}(z) := \|k\delta Sz + x_j\|$$

which is 1-Lipschitz ($k\delta < 1$ and S is 1-Lipschitz). Denote $a_{j,k} = \int_{\mathbb{S}^{n-1}} f_{j,k}\, d\mu$ where μ is the normalized Haar measure on $\mathbb{S}^{n-1}$. Then by concentration of measure results, we have that

$$\mu\big(z \in \mathbb{S}^{n-1}\colon |f_{j,k}(z) - a_{j,k}| \geq \delta\big) \leq K e^{-Cn\delta^2}$$

Let ν be the Haar measure on $O(n)$, and consider the event $A \subseteq O(n)$ defined by

$$A = \left\{ U \in O(n)\colon |f_{j,k}(Uz_i) - a_{j,k}| \geq \delta \text{ for some } 1 \leq i \leq M, 1 \leq j \leq N, 1 \leq k \leq \frac{1}{\delta} \right\}$$

Then

$$\begin{aligned} \nu(A) &= \sum_{i=1}^{M} \nu\left(U \in O(n)\colon |f_{j,k}(Uz_i) - a_{j,k}| \geq \delta \text{ for some } j \leq N, k \leq \frac{1}{\delta} \right) \\ &= \sum_{i=1}^{M} \mu\left(z \in \mathbb{S}^{n-1}\colon |f_{j,k}(z) - a_{j,k}| \geq \delta \text{ for some} j \leq N, k \leq \frac{1}{\delta} \right) \\ &\leq \frac{1}{\delta} MNke^{-cn\delta^2} \end{aligned}$$

which for n large enough is less than 1.

This means that there exists $U \in O(n)$ such that if we define $T := SU\colon l_2^n \to X$, then for all $i \leq M$, $j \leq N$ and $k \leq \frac{1}{\delta}$ we have that

$$|\|k\delta T z_i + x_j\| - a_{j,k}| \leq \delta$$

so that $|\|k\delta T z_i + x_j\| - \|-k\delta T z_i + x_j\|| \leq 2\delta$.

For $z \in \mathbb{S}^{n-1}$, we have that using nets and the 1-Lipschitz property of $f_{j,k}$,

$$|\,\|x_j + k\delta T z\| - \|x_j - k\delta T z\|| \leq 4\delta$$

Now take $\|z\| \leq 1$. Since the above holds for $k \leq \frac{1}{\delta}$, we'll use k for which $|\|z\| - k\delta| \leq \delta$. Then

$$\left| \left\| x_j + k\delta T\left(\frac{z}{\|z\|}\right) \right\| - \left\| x_j - k\delta T\left(\frac{z}{\|z\|}\right) \right\| \right| \leq 4\delta$$

and

$$|\,\|x_j + Tz\| - \|x_j - Tz\|\,| \leq 6\delta$$

Finally using the δ-net, we have that for all $x \in B_E$ and for all $z \in B_{l_2^n}$,

$$|\,\|x + Tz\| - \|x - Tz\|\,| \leq 8\delta$$

Define $F = Tl_2^n \subseteq X$. We know that for all $e \in E$, $f \in F$, that $|\,\|e + f\| - \|e - f\|\,| \leq 8\delta \max\{\|e\|, \|f\|\}$. (The previous statement with homogeneity). By triangle inequality, we know that

$$\max\{\|e\|, \|f\|\} \leq \frac{\|e + f\| + \|e - f\|}{2}$$

thus $\|e + f\| - \|e - f\| \leq 4\delta(\|e + f\| + \|e - f\|)$ and

$$\|e + f\| \leq \frac{1 + 4\delta}{1 - 4\delta} \|e - f\|$$

for all $e \in E$, $f \in F$. Note $E + F \subseteq X$.

Now for $Y = E \oplus l_2^n$, define $\|(x, z)\|_Y = \|x + Tz\|$. Triangle inequality and homogeneity follow easily. To show that if the norm is 0 that both x, z are 0, we note that

$$\begin{aligned}\|x+Tz\| &= \frac{1}{2}(\|x+Tz\| + \|x+Tz\|) \\ &\geq \frac{1-4\delta}{2(1+4\delta)}(\|x+Tz\| + \|x-Tz\|) \\ &\geq \frac{1-4\delta}{1+4\delta}\max(\|x\|, \|Tz\|)\end{aligned}$$

so that if $\|x+Tz\| = 0$, both $\|x\|$ and $\|Tz\| = 0$.

Now we set $R: Y \to X$ with $R(x, z) = x + Tz$ and we see that all the properties are satisfied. The main headache was the last property, which we used a concentration of measure argument to obtain. □

Theorem 69. *Let X be an infinite dimensional Banach space, and $\lambda > 1$, such that for all finite dimensional subspaces $E \subseteq X$, there exists a projection P: $X \to E$ (onto) with $\|P\| \leq \lambda$. Then for all finite dimensional subspaces $F \subseteq X$, $d_F \leq 4\lambda^2$.*

Note that starting from the statement of Lindenstrauss-Tzafriri, knowing that all closed subspaces are complemented, Theorem 64 shows us that all finite dimensional subspaces are complemented with a projection with a norm $\leq \lambda$ for some λ, satisfying the conditions of this theorem. The conclusion of this theorem combined with Lemma 67 proves the theorem.

Proof. Fix $E \subseteq X$ and $\dim E = n$. By previous theorem (68) there exists a norm $\|\cdot\|_Y$ on $E \oplus l_2^n$ such that

1. $(E \oplus l_2^n, \|\cdot\|_Y)$ is isometric to a subspace of X.
2. $\|(x, 0)\|_Y = \|x\|$
3. $(1-\varepsilon)\|z\| \leq \|(0, z)\| \leq \|z\|$
4. $(1-\varepsilon)\|(x, -z)\|_Y \leq \|(x, z)\|_Y \leq (1+\varepsilon)\|(x, -z)\|_Y$

Note $\|(x, z)\|_Y \geq (1-2\varepsilon)\max\{\|x\|, \|z\|\}$ since

$$\begin{aligned}\|(x,z)\|_Y &\geq \frac{1-\varepsilon}{2}(\|(x,z)\|_Y + \|(x,-z)\|_Y) \\ &\geq (1-\varepsilon)\|(x,0)\|_Y \\ &= (1-\varepsilon)\|x\|\end{aligned}$$

and

$$\begin{aligned}\|(x,z)\|_Y &\geq \frac{1-\varepsilon}{2}(\|(x,z)\|_Y + \|(-x,z)\|_Y) \\ &\geq (1-\varepsilon)\|(0,z)\|_Y \\ &\geq (1-\varepsilon)^2\|z\|\end{aligned}$$

We will be using this often below.

Define $\theta = \sqrt{d_E}$. There exists a linear $S: E \to l_2^n$ with $\frac{1}{\theta}\|x\| \leq \|Sx\| \leq \theta\|x\|$. Consider the subspace

$$Z = \{(x, Sx): x \in E\} \subseteq Y$$

There exists a projection $P\colon Y \to Z$ (onto) with $\|P\| \leq \lambda$ by assumption. Define $T\colon E \to l_2^n$ by the second coordinate of the projection, i.e. if $P(x,z) = (a, Sa)$, then $Tx = Sa$. In other words, if we define $Q\colon Y \to l_2^n$ by $Q(x,z) := z$, then $T = QP$. Note $P(x,0) = (S^{-1}Tx, Tx)$.

We know that

$$\|P(x,0)\| \leq \lambda \|(x,0)\| = \lambda \|x\|$$

and

$$\|P(x,0)\| = \|(S^{-1}Tx, Tx)\| \geq (1-2\varepsilon)\|Tx\|$$

This implies that $\|T\| \leq \frac{\lambda}{1-2\varepsilon}$.

Let $V\colon E \to l_2^{2n} = l_2^n \oplus l_2^n$ defined by $Vx = (\lambda Sx, \theta Tx)$. Then

$$\begin{aligned} \|Vx\|^2 &= \lambda^2 \|Sx\|^2 + \theta^2 \|Tx\|^2 \\ &\leq \lambda^2 \theta^2 \|x\|^2 + \frac{\theta^2 \lambda^2}{(1-2\varepsilon)^2} \|x\|^2 \\ &\leq \lambda^2 \theta^2 (1+10\varepsilon)^2 \|x\|^2 \end{aligned}$$

and thus $\|V\| \leq \theta\lambda(1+10\varepsilon)$.

To bound the norm of the inverse, note

$$\begin{aligned} P(0, Sx) &= P(x, Sx) - P(x, 0) \\ &= (x, Sx) - (S^{-1}Tx, Tx) \\ &= (x - S^{-1}Tx, Sx - Tx) \end{aligned}$$

so that

$$\|P(0,Sx)\| \geq (1-2\varepsilon)\|x - S^{-1}Tx\|$$

$$\|P(0,Sx)\| \leq \lambda \|(0,Sx)\| \leq \lambda \|Sx\|$$

which shows that

$$\|x - S^{-1}Tx\| \leq \frac{\lambda}{1-2\varepsilon} \|Sx\|$$

Then

$$\begin{aligned} \|x\| &\leq \|S^{-1}Tx\| + \|x - S^{-1}Tx\| \\ &\leq \theta \|Tx\| + \frac{\lambda}{1-2\varepsilon} \|Sx\| \\ &\leq \frac{\sqrt{2}}{1-2\varepsilon} \left(\theta^2 \|Tx\|^2 + \frac{\lambda^2}{(1-2\varepsilon)^2} \|Sx\|^2 \right)^{1/2} = \frac{\sqrt{2}}{1-2\varepsilon} \|Vx\| \end{aligned}$$

which shows that $\|V^{-1}\| \leq \frac{\sqrt{2}}{1-2\varepsilon}$, and

$$\theta^2 = d_E \leq \|V\| \|V^{-1}\| \leq (1+10\varepsilon)\sqrt{2}\,\theta\lambda \frac{\sqrt{2}}{1-2\varepsilon} \leq (1+100\varepsilon) 2\theta\lambda$$

and since $\theta = \sqrt{d_E}$, this shows that $d_E \leq 4\lambda^2$. (ε arbitrary)

$\square$

Week 11 (11/16/2010)

Nonlinear Dvoretsky

Let $(X, d_X), (Y, d_Y)$ be metric spaces. We use the notation $X \overset{D}{\hookrightarrow} Y$ to say that there exists $f: X \to Y$ and $\lambda > 0$ such that

$$\lambda d_X(x, y) \leq d_Y(f(x), f(y)) \leq D\lambda d_X(x, y)$$

If Y is a Banach space, we can assume that $\lambda = 1$ by replacing f by λf.

Then a funny way to state Dvorestky's Theorem: For all $k \in \mathbb{N}$, for all $D > 1$, there exists $n = n(k, D)$ such that any n dimensional Banach space X has a linear subspace $Y \subseteq X$ with $\dim Y \geq k$ and $Y \overset{D}{\hookrightarrow} l_2$. Our version of Dvoretsky's Theorem is the same but the notation $\hookrightarrow$ was used only for linear maps. The Radamacher differentiation theorem can be used to show that these two are equivalent.

Theorem 70. (Nonlinear Dvoretsky) *For all $\varepsilon \in (0, 1)$, any n-point metric space (X, d_X) has $Y \subseteq X$ such that*

1. $|Y| \geq n^{1-\varepsilon}$
2. $Y \overset{O(1/\varepsilon)}{\hookrightarrow} l_2$

This theorem is sharp, meaning that for all n, there exists an n-point metric space X_n such that for all $Y \subseteq X_n$, if $|Y| \geq n^{1-\varepsilon}$ then it has distortion $\geq C/\varepsilon$ in any embedding into l_2.

Bourgain's Embedding Theorem: $|X| = n$, $X \overset{100 \log n}{\hookrightarrow} l_2$. This is also sharp.

Notation: (X, d) metric space, then $B(x, r) = \{y \in X: d(x, y) \leq r\}$. Let $A \subseteq X$, $\operatorname{diam} A = \sup_{x, y \in A} d(x, y)$. If P is a partition of X, then

1. For $x \in X$, let $P(x)$ denote the unique set in P to which x belongs.
2. P is called Δ-bounded if $\operatorname{diam} P(x) \leq D$ for all $x \in X$

Definitions: A sequence of partitions of X $\{P_k\}_{k=0}^{\infty}$ is called a partition tree if

1. $P_0 = \{x\}$
2. P_k is a $8^{-k} \operatorname{diam}(X)$-bounded
3. P_{k+1} is a refinement of P_k, i.e. $P_{k+1}(x) \subseteq P_k(x)$

A distribution over partition trees $\{P_k\}_{k=0}^{\infty}$ is called **completely β-padded with exponent γ** if for all $x \in X$,

$$\Pr\Big[\forall k, B(x, \beta 8^{-k}) \subseteq P(x)\Big] \geq \frac{1}{n^{\gamma}}$$

where $n = |X|$.

Lemma 71. *If X admits a random partition tree completely β-padded with exponent γ then there exists $Y \subseteq X$ such that*

1. $|Y| \geq n^{1-\gamma}$

2. $Y \overset{8/\beta}{\hookrightarrow} l_2$

Proof. Let $\{P_k\}_{k=0}^{\infty}$ be the random partition tree. Define a random subset:

$$Y = \{x \in X, \forall k B(x, \beta 8^{-k}) \subseteq P_k(x)\}$$

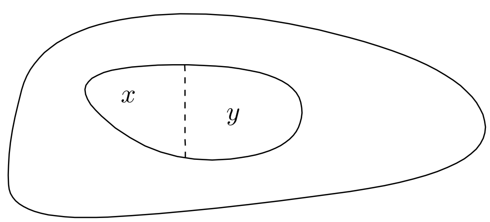
Then

$$\mathbb{E}|Y| = \mathbb{E}\left[\sum_{x \in X} \mathbf{1}_{\{x \in Y\}}\right] = \sum_{x \in X} \Pr(x \in Y) \geq \sum_{x \in X} \frac{1}{n^{\gamma}} = n^{1-\gamma}$$

Thus there exists Y with $|Y| \geq n^{1-\gamma}$. We will show that by construction, Y is isometric to l_2.

For all x, y distinct, let $K(x, y)$ be the largest $k \in \mathbb{N}$ such that $P_k(x) = P_k(y)$. Define $\rho(x, y) := 8^{-K(x,y)}$.

Observation: ρ is a (random) ultrametric on X (i.e. $\rho(x, y) \leq \max\{\rho(x, z), \rho(y, z)\}$). We can illustrate this with a diagram: Look at the moment when x, y stop being in the same component. Below the dotted line represents the separation between $P_{k+1}(x)$ and $P_{k+1}(y)$, and the region around is $P_k(x) = P_k(y)$.



Suppose $z \notin P_k(x)$. Then $\rho(x, z) = \rho(y, z)$ (whenever z separates from x is exactly when z separates from y), and furthermore $K(x, y) \geq K(x, z)$ and thus $\rho(x, y) \leq \rho(x, z)$. Suppose $z \in P_{k+1}(x)$. Then $K(x, y) = K(z, y)$, and $K(x, z) \geq K(y, z)$ so that $\max\{\rho(x, z), \rho(y, z)\} = \rho(y, z) = \rho(x, y)$. Swapping x, y shows the result for when $z \in P_{k+1}(y)$. Finally, if $z \in P_k(x)$ but not in $P_{k+1}(x)$ or $P_{k+1}(y)$, then $K(x, z) = K(y, z) = k+1$, and $K(x, y) = k$, so that $\rho(x, y) = k \leq k+1 = \max\{\rho(x, z) = \rho(y, z)\}$. This shows the observation.

Suppose $x, y \in Y$.

We note for $k \leq K(x, y)$, $x, y \in P_k(x) = P_k(y)$ so that $d(x, y) \leq \operatorname{diam}(P_k(x)) \leq 8^{-k}$ and thus

$$d(x, y) \leq 8^{-K(x,y)} = \rho(x, y)$$

Since $P_{K+1}(x) \neq P_{K+1}(y)$, and $x, y \in Y$, $B(x, \beta 8^{-K-1}) \cap B(y, \beta 8^{-K-1}) = \varnothing$ so that

$$d(x, y) \geq \beta 8^{-K(x,y)-1} = \frac{\beta}{8} \rho(x, y)$$

Thus, for all $x, y \in Y$, we have

$$\frac{\beta}{8} \rho(x, y) \leq d(x, y) \leq \rho(x, y)$$

We have shown that $(Y, d_X) \overset{8/\beta}{\hookrightarrow} (Y, \rho)$, where ρ is an ultrametric. It turns out that every finite ultrametric is isometric to a subset of a Hilbert space, and this shows the result. We will prove this in the following lemma. $\square$

Lemma 72. *Every finite ultrametric is isometric to a subset of a Hilbert space.*

Proof. (By induction on $n = |X|$). We will prove that there exists a function $f: X \to H$ Hilbert space such that

1. $\rho(x, y) = \|f(x) - f(y)\|$ for all $x, y \in X$
2. $\|f(x)\| = \frac{\mathrm{diam}(X)}{\sqrt{2}}$ for all $x \in X$

We say that $x \sim y$ if $\rho(x, y) < \mathrm{diam}(X)$. This is an equivalence relation because ρ is an ultrametric: Clearly $x \sim x$ and $x \sim y \Longleftrightarrow y \sim x$. Suppose $x \sim y$ and $y \sim z$. Then since $\rho(x, z) \leq \max\{\rho(x, y), \rho(y, z)\} \leq \mathrm{diam}(X)$, we have $x \sim z$, so we have transitivity of $\sim$.

Let $A_1..., A_k$ be the equivalence classes. $|A_i| < |X| = n$. By induction, there exists $f_i: A_i \to H_i$ such that

1. $\rho(x, y) = \|f_i(x) - f_i(y)\|$ for all $x, y \in A_i$
2. $\|f_i(x)\| = \frac{\mathrm{diam}(A_i)}{\sqrt{2}}$ for all $x \in X$.

Define $f: X \to \left(\bigoplus_{i=1}^{k} H_i\right) \oplus l_2^k$. For $x \in X$, we have that $x \in A_i$ for some i, and we set

$$f(x) = \sqrt{\frac{\mathrm{diam}(X)^2 - \mathrm{diam}(A_i)^2}{2}} e_i + f_i(x)$$

so that

$$\|f(x)\| = \sqrt{\frac{\mathrm{diam}(X)^2 - \mathrm{diam}(A_i)^2}{2} + \frac{\mathrm{diam}(A_i)^2}{2}} = \frac{\mathrm{diam}(X)}{2}$$

If $x, y \in A_i$, then

$$\|f(x) - f(y)\| = \|f_i(x) - f_i(y)\| = \rho(x, y)$$

If $x \in A_i$ and $y \in A_j$ for $i \neq j$, we have

$$\|f(x) - f(y)\| = \sqrt{\|f(x)\|^2 + \|f(y)\|^2} = \sqrt{\frac{\mathrm{diam}(X_i)^2}{2} + \frac{\mathrm{diam}(X_i)^2}{2}} = \mathrm{diam}(X) = \rho(x, y)$$

□

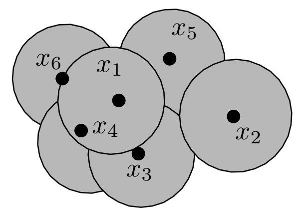
Now we just need a way to generate a suitable distribution over random partition trees.

Lemma 73. *Let X be an n-point metric space, $\Delta > 0$. Then there exists a distribution over random Δ-bounded partitions P such that for all $0 < t < \frac{\Delta}{8}$ and for all $x \in X$*

$$\Pr(B(x, t) \subseteq P(x)) \geq \left(\frac{|B(x, \Delta/8)|}{|B(x, \Delta)|}\right)^{8t/\Delta}$$

Proof. Let $X = \{x_1, ..., x_n\}$. We will be constructing the partition randomly out of metric balls of diameter R, which will also be randomly chosen. The method we will use will depend highly on the order that we process the elements of X. Thus, take a random ordering, i.e. take $\pi \in S_n$ to be a random permutation chosen uniformly amongst the $n!$ permutations. Also, let R be uniformly distributed over $[\Delta/4, \Delta/2]$.

Take $C_1 = B(X_{\pi(1)}, R)$, and inductively given C_1, ..., C_{j-1}, take $B(X_{\pi(j)}) \setminus \bigcup_{i=1}^{j-1} C_i$. Sample partition of six points:



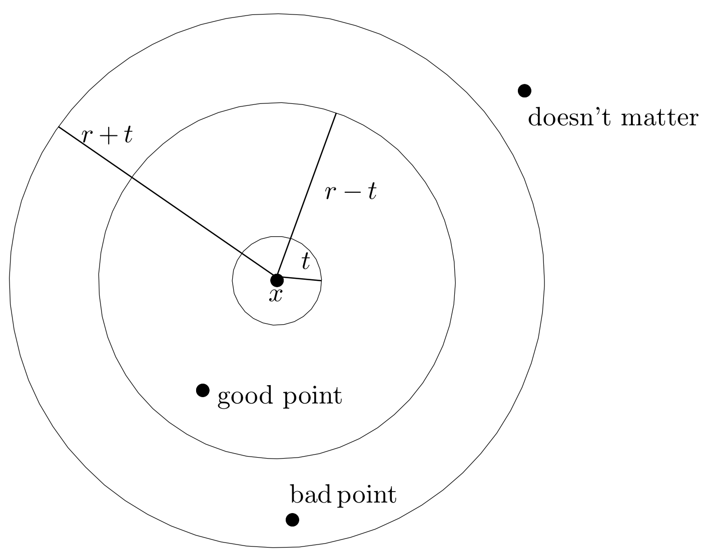
(Should imagine this for more points, but note x_1, x_4 are in the same partition. Also note that this is highly dependent on the order of the points; for instance, it would be an unlucky situation of all points were within R of the first point $x_{\pi(1)}$ that we choose)

The rest is an exercise in probability.

Claim: For all $R \in [\Delta/4, \Delta/2]$, and for all $x \in X$, and $t < \Delta/8$,

$$\Pr[B(x,t) \subseteq P(x) \colon R = r] \geq \frac{|B(x, r-t)|}{|B(x, r+t)|}$$

Note the diagram below:



The "good points", i.e. points inside $B(x,\, r - t)$, are points y such that $B(x,\, t) \subset B(y,\, r)$ (triangle inequality). On the other hand, "bad points", i.e. points inside $B(x,\, r + t)$ but not in $B(x,\, r - t)$, are points z such that $B(x,t) \cap B(z,r) \neq \varnothing$. If in the random ordering a bad point is processed before a good point, then part of $B(x,t)$ will be in $B(z,r)$, so that $B(x,t)$ will not be contained in $P(x)$ of the resulting partition. Points outside $B(x,r+t)$ do not matter since for such $w \notin B(x,r+t)$, $B(w,r) \cap B(x,t) = \varnothing$.

Thus, the probability that for a particular x, that $B(x,\, t) \subseteq P(x)$ is the probability that the first point in the random order that lies in $B(x,\, r + t)$ also happens to be in $B(x,\, r - t)$, and this occurs with probability $\frac{|B(x,\, r-t)|}{|B(x,\, r+t)|}$. This shows the claim.

We are conditioning on $R = r$, so we can then compute the full probability:

$$\begin{aligned}
\Pr(B(x,t) \subseteq P(x)) &= \frac{1}{\Delta/4} \int_{\Delta/4}^{\Delta/2} \Pr(B(x,t) \subseteq P(x) \,|\, R = r)\, dr \\
\text{(By the claim)} &\geq \frac{1}{\Delta/4} \int_{\Delta/4}^{\Delta/2} \frac{|B(x, r-t)|}{|B(x, r+t)|}\, dr
\end{aligned}$$

Let $h(r) = \log |B(x,r)|$. Then

$$\begin{aligned}
&= \frac{1}{\Delta/4} \int_{\Delta/4}^{\Delta/2} e^{h(r-t)+h(r+t)}\, dr \\
\text{(Jensen)} &\geq \exp\left\{ \frac{1}{\Delta/4} \int_{\Delta/4}^{\Delta/2} h(r-t) + h(r+t)\, dr \right\} \\
&= \exp\left\{ \frac{1}{\Delta/4} \left[\int_{\Delta/4-t}^{\Delta/2-t} h(r) - \int_{\Delta/4+t}^{\Delta/2+t} h(r)\, dr \right] \right\} \\
\text{monotonicity of } h(r) &\geq \exp\left\{ \frac{2t}{\Delta/4} [\, h(\Delta/4 - t) - h(\Delta/2 + t)] \right\} \\
(t < \Delta/8) &\geq \exp\left\{ \frac{8t}{\Delta} (h(\Delta/8) - h(\Delta)) \right\} = \left(\frac{|B(x, \Delta/8)|}{|B(x, \Delta)|} \right)^{8t/\Delta}
\end{aligned}$$

$\square$

Now we can complete the proof of the nonlinear Dvoretsky theorem:

Proof. (of Theorem 70) Without loss of generality, $\operatorname{diam}(X) = 1$. $P_0 = \{X\}$. For all $k \geq 1$, let P_k be a random partition applying the previous lemma with $\Delta = 8^{-k}$, with $\{P_k\}$ independent of one another. This is not yet a partition tree because the refinement property is not satisfied. Let Q_k be the common refinement of $P_1 \ldots, P_k$, so that $Q_{k+1}(x) = Q_k(x) \cap P_{k+1}(x)$.

Note by the common refinment property,

$$\{\forall k, B(x, \beta 8^{-k}) \subseteq P_k(x)\} \subseteq \{\forall k, B(x, \beta 8^{-k}) \subseteq Q_k(x)\}$$

This is because $B(x, \beta 8^{-k}) \subseteq P_l(x)$ for all $l \leq k$.

Now we have

$$\begin{aligned}
\Pr(\forall k\, B(x, \beta 8^{-k}) \subseteq Q_k(x)) &\geq \Pr(\forall k B(x, \beta 8^{-k}) \subseteq P_k(x)) \\
&= \prod_{k=0}^{\infty} \Pr(B(x, \beta 8^{-k}) \subseteq P_k(x)) \\
&\geq \prod_{k=0}^{\infty} \left(\frac{|B(x, 8^{-k-1})|}{|B(x, 8^{-k})|} \right)^{8\beta} \\
\text{(telescoping)} &= \frac{1}{n^{8\beta}}
\end{aligned}$$

Above we used the lemma with $t = \beta 8^{-k} \leq \frac{\Delta}{8} = \frac{8^{-k}}{8}$, i.e. $\beta \leq \frac{1}{8}$. Thus, for all $\beta \leq \frac{1}{8}$, X admits a random partition tree β-padded with exponent 8β, and thus by Lemma 71 there exists $Y \subseteq X$ with $|Y| \geq n^{1-8\beta}$ and $Y \overset{8/\beta}{\hookrightarrow} l_2$ $\square$

Sharpness of Nonlinear Dvoretsky

For the sharpness, recall we are finding an n-point metric space for which large subsets $|Y| \geq n^{1-\varepsilon}$ will have distortion $\geq C/\varepsilon$ in any embedding into l_2. We will be holding ε fixed, though a stronger result allows us to prove the result with ε that depends on n. The proof will be by a counting argument to show the existence of a random n-point metric space.

Some terminology: $G=(V,E)$ is an n-vertex graph. The **girth** of G is the length of the shortest cycle in G. For $A\subseteq V$, let E_A be the set of edges involving just vertices A, denote $e_A := |E_A|$ and the **density** of A is defined to be $\frac{e_A}{\binom{|A|}{2}}$.

Let X be a Banach space, $D>1$. Denote $R_X(D,n)$ be the largest integer k such that any n-point metric space X_n has $Y\subseteq X_n$ with $|Y|\geq k$, $Y \overset{D}{\hookrightarrow} X$. We will look at $R_{l_2}(D,n)$.

Lemma 74. *Assume that there exists an n-vertex graph $G=(V,E)$ of girth g such that every subset of V of size $\geq s$ has density $\geq q$. Then for all h-dimensional Banach spaces X, for all $1\leq D\leq g-1$,*

$$R_X(D,n)\leq \max\left\{s, \frac{2}{q}\left(h\log_2\left(\frac{14Dg}{g-D-1}\right)+\log_2\left(\binom{n}{s}\right)\right)\right\}$$

Lemma 75. *For all integers $g\geq 4$, there exists an arbitrarily large n-vertex graph with girth $\geq g$ and every subset $\subseteq V$ of size $n^{1-\frac{1}{8g}}$ has density $\geq n^{-1+\frac{1}{2g}}$.*

These two lemmas imply sharpness (i.e. $R_{l_2}(C/\varepsilon,n)\lesssim n^{1-\varepsilon}$). Let $D>1$ and $k=R_{l_2}(D,n)$. By Johnson-Lindenstrauss, every n-point metric space has a subset of size k that embeds with distortion $\leq 2D$ into l_2^h where $h\leq C\log k$. Take $g=\lceil D+2\rceil$, then the two lemmas apply with $s=n^{1-\frac{1}{8g}}$ and $q=n^{-1+\frac{1}{2g}}$ and $h=C\log k$.

Then noting $g=O(d)$,

$$k\leq C\max\left\{n^{1-\frac{c}{D}}, n^{1-\frac{c}{D}}\left(\log k\log D+\frac{\log n}{D}\right)\right\}$$

so that $k\lesssim Cn^{1-\frac{c}{D}}\log n\,\log D$. Plugging in $D=C/\varepsilon$ we have that $R_{l_2}(C/\varepsilon,n)\lesssim n^{1-\varepsilon}$ (other terms are lower order terms)

Proof. (Lemma 74) For all $H\subseteq E$ define a metric ρ_H on V

$$\rho(u,v)=\min\{g-1, d_H(u,v)\}$$

where d_H is the shortest path metric in (V,H). Let $k=R_X(D,n)$ and we may assume that $k\geq s$, otherwise the lemma holds already. For all $H\subseteq E$, there exists $A_H\subseteq V$, $|A_H|=k$ and $(A_H,\rho_H)\overset{D}{\hookrightarrow} X$ by definition of $R_X(D,n)=k$. There are $2^{|E|}$ such H and only $\binom{n}{k}$ possible A_H, so by pigeonhole, there exists a family $\mathcal{H}$ of subsets of E and a k-point subset $A\subseteq V$ such that

$$\forall H\in\mathcal{H}, A_H=A, |\mathcal{H}|\geq\frac{2^{|E|}}{\binom{n}{k}}$$

Let $H_1,H_2\in\mathcal{H}$. Say that $H_1\sim H_2$ if $H_1\cap E_A=H_2\cap E_A$. The number of equivalent H_1 is bounded above by $2^{|E|-e_A}$, which means that by pigeonhole there exists at least $\frac{|\mathcal{H}|}{2^{|E|-e_A}}\geq\frac{2^{e_A}}{\binom{n}{k}}$ inequivalent elements in $\mathcal{H}$.

Summarizing above: There exists $A \subseteq V$ with $|A| = k$ and $H_1, \ldots, H_m \subseteq E$ such that

1. $M \geq \frac{2^{e_A}}{\binom{n}{k}}$

2. $H_i \cap E_A \neq H_j \cap E_A$ for all $i \neq j$

3. $(A, \rho_{H_i}) \overset{D}{\hookrightarrow} X$

If $H \subseteq E$ is one of the H_i, there exists $f_H: A \to X$ such that

$$\frac{1}{D} \rho_H(u, v) \leq \| f_H(u) - f_H(v) \| \leq \rho_H(u, v) \text{ for all } u, v \in A$$

Since ρ_H takes values in $\{0, \ldots, g - 1\}$, by translation without loss of generality $f_H(A) \subseteq B(0, g)$. Let $\mathcal{N}$ be a δ-net in $B(0, g)$. For all $u \in A$, define $\phi_H(u) \in \mathcal{N}$ to be some element of X at distance $\leq \delta$ from $f_H(u)$. So $\phi_H: A \to \mathcal{N}$.

Key observation: If $i \neq j$, then $\phi_{H_i} \neq \phi_{H_j}$.

To show this, if $i \neq j$, since H_i and H_j are not equivalent, there exists $u, v \in A$ with $u \neq v$ such that the edge $\{u, v\}$ is in H_i but not H_j. Then we note $\rho_{H_i}(u, v) = 1$ (since $\{u, v\}$ is an edge in H_i)

$$\rho_{H_j}(u, v) = \min \{g - 1, d_{H_j}(u, v)\} = g - 1$$

noting $d_{H_j}(u, v) \geq g - 1$ or else a shorter cycle exists ($\{u, v\} \in H_i$ and the path from $u \to v$ in H_j), contradicting the girth of G being g.

Now assume towards a contradiction that $\phi_{H_i}(u) = \phi_{H_j}(u)$ and $\phi_{H_i}(v) = \phi_{H_j}(v)$.

Using the f_H above, we have that

$$\begin{aligned}
\frac{g-1}{D} = \frac{\rho_{H_j}(u, v)}{D} &\leq \| f_{H_j}(u) - f_{H_j}(v) \| \\
&\leq 2\delta + \| \phi_{H_j}(u) - \phi_{H_j}(v) \| \\
&= 2\delta + \| \phi_{H_i}(u) - \phi_{H_i}(v) \| \\
&\leq 4\delta + \| f_{H_i}(u) - f_{H_i}(v) \| \\
&\leq 4\delta + \rho_{H_i}(u, v) \\
&= 4\delta + 1
\end{aligned}$$

Thus $\frac{g-1}{D} \leq 4\delta + 1$, which is a contradiction if $\delta = \left(\frac{g-1}{D} - 1 \right) \frac{1}{5}$.

Note that the number of mappings ϕ: $A \to \mathcal{N}$ is $|\mathcal{N}|^k \leq \left(\frac{2g}{\delta} \right)^{kh}$ (bound on size of δ-net of $B(0, g)$ contained in a h-dimensional space), and so

$$\frac{2^{q\binom{k}{2}}}{\binom{n}{k}} \leq \frac{2^{e_A}}{\binom{n}{k}} \leq M \leq |\mathcal{N}|^k \leq \left(\frac{2g}{\delta} \right)^{kh}$$

Above we used the desnity $\frac{e_A}{\binom{k}{2}} \geq q$ where A is k-dimensional, with $k \geq s$ by assumption. Now using $\binom{n}{k} \leq \left(\frac{en}{k} \right)^k$ and plugging in $\delta = \left(\frac{g-1}{D} - 1 \right) \frac{1}{5}$, we have (using $k \geq s$ when k is in the denominator)

$$2^{qk(k+1)/2} \leq \left(\frac{10\, g\, D}{(g - 1 - D)} \right)^{kh} \left(\frac{en}{s} \right)^k$$

Taking log of both sides,

$$q\frac{k^2}{2} \leq kh\log_2\left(\frac{10Dg}{g-D-1}\right) + k\log_2\left(\frac{en}{s}\right)$$

or

$$k \leq \frac{2}{q}\left(h\log_2\left(\frac{10Dg}{g-D-1}\right) + \log_2\left(\frac{en}{s}\right)\right)$$

□

To finish everything, we show the proof of the second Lemma:

Proof. (of Lemma 75) To finish. □

Week 12 (11/30/2010)

Embeddings in l_∞ and l_2 (Alon, Milman, Talagrand)

Student Presentation: Lukas Koehler, Evan Chou

We look at a result by Alon, Milman, with an easier proof by Talagrand.

Consider x_1, ..., x_n normalized vectors in a Banach space X, let ε_i be i.i.d random variables distributed uniformly on $\{\pm 1\}$, and set

$$M_n = \mathbb{E}_{\varepsilon_i=\pm 1}\left[\left\| \sum_{i=1}^n \varepsilon_i x_i \right\|\right] = \mathbb{E}\left[\sup_{x^* \in B_{X^*}}\left| \sum_{i=1}^n \varepsilon_i x^*(x_i)\right|\right]$$

$$\omega_n = \sup\left\{ \sum_{i=1}^n |x^*(x_i)|, x^* \in B_{X^*}\right\}$$

Note immediately that $M_n \leq \omega_n$ (take absolute value inside the sum).

We can also obtain a reverse inequality: For any $x^* \in B_{X^*}$,

$$M_n \geq \mathbb{E}\left[\left| \sum_{i=1}^n \varepsilon_i x^*(x_i)\right|\right]$$

Kintchine's inequality:

$$A_p\|a\|_2 \leq \left(\mathbb{E}_{\varepsilon\in\{\pm 1\}}\left| \sum_{i=1}^n \varepsilon_i a_i\right|^p\right)^{1/p} \leq B_p\|a\|_2$$

For the lower bound with $p=1$ we can use $A_1 = \frac{1}{\sqrt{2}}$. Thus

$$\sqrt{2}\, M_n \geq \left(\sum_{i=1}^n |x^*(x_i)|^2\right)^{1/2}$$

and by Cauchy Schwarz,

$$\omega_n \leq \sqrt{2n}\, M_n$$

Then we have the following result:

Theorem 76. *There exists a subset A of $\{1, ..., n\}$ of cardinality k at least $\frac{n}{64\omega_n}$ such that $(x_i)_{i\in A}$ is $8M_n$ isomorphic to the natural basis of l_∞^k, i.e. there exist two constants C_1, C_2 such that*

$$C_1\|a\|_\infty \leq \left\| \sum_{i\in A} a_i x_i \right\| \leq C_2 \|a\|_\infty \text{ for all } a \in l_\infty(A)$$

with $\frac{C_2}{C_1} = 8\, M_n$.

The proof by Alon-Milman has $k \geq \frac{\sqrt{n}}{2^7 M_n}$ and $16\, M_n$ isomorphic. Note that we have a smaller isomorphism constant here and $\frac{n}{64\omega_n} \geq \frac{n}{64\sqrt{2n}\, M_n} = \frac{\sqrt{n}}{2^{6.5} M_n}$, and the proof will be simpler.

The Upper Bound:

We will be generating random subsets that achieve an upper bound. Note

$$\left\| \sum_{i\in A} a_i x_i \right\| = \sup_{x^*\in B_{X^*}} \sum_{i\in A} a_i\, x^*(x_i) \leq \|a\|_\infty \sup_{x^*\in B_{X^*}} \sum_{i\in A} |x^*(x_i)|$$

If we use $A = \{1, ..., n\}$, then the bound becomes $\omega_n\, \|a\|_\infty$. The main point is that we can find a subset which reduces the bound to $4M_n\, \|a\|_\infty$.

To generate subsets, we will use indicators δ_i which are i.i.d random variables uniform on $\{0, 1\}$.

Proposition 77. *Consider $\delta = M_n/\omega_n$, and independent r.v. δ_i such that $P(\delta_i = 1) = \delta$ and $P(\delta_i = 0) = 1 - \delta$. Then*

$$\mathbb{E} \sup_{x^*\in B_{X^*}} \sum_{i=1}^n \delta_i |x^*(x_i)| \leq 3\, M_n$$

There are two main ingredients: Note M_n has no absolute value sign with $x^*(x_i)$, we first look to put an absolute value there, where it turns out we only lose a factor of 2:

Lemma 78.

$$\mathbb{E} \sup_{x^*\in B_{X^*}} \left| \sum_{i=1}^n \varepsilon_i\, |x^*(x_i)| \right| \leq 2\mathbb{E} \sup_{x^*\in B_{X^*}} \left| \sum_{i=1}^n \varepsilon_i x^*(x_i) \right| = 2M_n$$

The second ingredient is that if we start removing indices, we still have the bound:

Lemma 79. *If we remove the j-th index, then*

$$\mathbb{E} \sup_{x^*\in B_{X^*}} \left| \sum_{i\neq j} \varepsilon_i\, |x^*(x_i)| \right| \leq 2\tilde{M}_n^{(j)} \leq 2\, M_n$$

where $\tilde{M}_n^{(j)} = \mathbb{E}\sup_{x^*\in B_{X^*}}\left|\sum_{i\neq j}\varepsilon_i\,|x^*(x_i)|\right|$. *(The first inequality is just Lemma 3)*

For the first Lemma, we note that we can replace B_{X^*} by an ε-net $\mathcal{N}_\varepsilon$, only affecting the inequality by an addition of ε, which is arbitrarily small (usual approximation argument). Thus we may assume that the supremum is over a finite set (so that the supremum is achieved)

Secondly, we note $x^*(x_i)_{i=1}^n$ is just a finite sequence, so the result reduces to proving:

Lemma 80. (Talagrand) *Let T be a finite subset of $\mathbb{R}^n$, and write $t=(t_1,\ldots,t_n)\in T$. Let ε_k be i.i.d* unif$\{\pm 1\}$. *Then*

$$\mathbb{E}\left[\sup_{t\in T}\left|\sum_{k=1}^n \varepsilon_k\,|t_k|\right|\right] \le 2\,\mathbb{E}\left[\sup_{t\in T}\left|\sum_{k=1}^n \varepsilon_k\,t_k\right|\right]$$

$(T=\{|x^*(x_i)|_{i=1}^n, x^*\in\mathcal{N}_\varepsilon)$

Proof. We will show that

$$\mathbb{E}\left[\sup_{t\in T}\left(\sum_{k=1}^n \varepsilon_k\,|t_k|\right)^+\right] \le \mathbb{E}\left[\sup_{t\in T}\left(\sum_{k=1}^n \varepsilon_k\,t_k\right)^+\right]$$

where $x^+=\max\{0,x\}$, and the result follows from noting that since $|x|=x^++x^-$ where $x^-=\min\{0,x\}$,

$$\sup_{t\in T}\left|\sum_{k=1}^n \varepsilon_k\,|t_k|\right| \le \sup_{t\in T}\left(\sum_{k=1}^n \varepsilon_k\,|t_k|\right)^+ + \sup_{t\in T}\left(\sum_{k=1}^n \varepsilon_k\,|t_k|\right)^-$$

and noting that $\sup_{t\in T}\left(\sum\varepsilon_k\,|t_k|\right)^-$ has the same distribution as $\sup_{t\in T}\left(\sum\varepsilon_k\,|t_k|\right)^+$ (flip all signs), we have that

$$\mathbb{E}\left[\sup_{t\in T}\left|\sum_{k=1}^n \varepsilon_k\,|t_k|\right|\right] \le 2\mathbb{E}\left[\sup_{t\in T}\left(\sum_{k=1}^n \varepsilon_k\,|t_k|\right)^+\right] \le 2\mathbb{E}\left[\sup_{t\in T}\left(\sum_{k=1}^n \varepsilon_k\,t_k\right)^+\right] \le 2\,\mathbb{E}\left[\sup_{t\in T}\left|\sum_{k=1}^n \varepsilon_k\,t_k\right|\right]$$

We will show the result by iteration, that replacing t_1 by $|t_1|$ decreases the expectation. If we condition on $\varepsilon_2,\ldots,\varepsilon_n$, then we can group $\left(t_1, \sum_{k=2}^N\varepsilon_k t_k\right)\in T$, and we are reduced to showing that

$$\mathbb{E}\left[\sup_{t\in T}(\varepsilon_1|t_1|+t_2)^+\right] \le \mathbb{E}\left[\sup_{t\in T}(\varepsilon_1 t_1+t_2)^+\right]$$

or

$$\frac{1}{2}\sup_t\,(|t_1|+t_2)^+ + \frac{1}{2}\sup_t\,(-|t_1|+t_2)^+ \le \frac{1}{2}\sup_t\,(t_1+t_2)^+ + \frac{1}{2}\sup_t\,(-t_1+t_2)^+$$

Let r maximize $|t_1|+t_2$ $(\varepsilon_1=1)$, and for $\varepsilon_1=-1$, let s maximize $-|s_1|+s_2$ $(\varepsilon_1=-1)$, so that the left hand side becomes

$$\frac{1}{2}(|r_1|+r_2)^+ + \frac{1}{2}(-|s_1|+s_2)^+$$

Now we look at different cases:

- $r_1 > 0, s_1 > 0$

$$\begin{aligned}\frac{1}{2}(|r_1| + r_2)^+ + \frac{1}{2}(-|s_1| + s_2)^+ &= \frac{1}{2}(r_1 + r_2)^+ + \frac{1}{2}(-s_1 + s_2)^+ \\ &\leq \frac{1}{2}\sup_t (t_1 + t_2)^+ + \frac{1}{2}\sup_t (-t_1 + t_2)^+\end{aligned}$$

- $r_1 < 0, s_1 < 0$. Here r and s swap roles.

$$\begin{aligned}\frac{1}{2}(|r_1| + r_2)^+ + \frac{1}{2}(-|s_1| + s_2)^+ &= \frac{1}{2}(-r_1 + r_2)^+ + \frac{1}{2}(s_1 + s_2)^+ \\ &\leq \frac{1}{2}\sup_t (-t_1 + t_2)^+ + \frac{1}{2}\sup_t (t_1 + t_2)^+\end{aligned}$$

- $r_1 < 0, s_1 > 0$. We get improvement by flipping the sign associated with the negative number r_1.

$$\begin{aligned}\frac{1}{2}(|r_1| + r_2)^+ + \frac{1}{2}(-|s_1| + s_2)^+ &= \frac{1}{2}(-r_1 + r_2)^+ + \frac{1}{2}(-s_1 + s_2)^+ \\ &\leq \frac{1}{2}(r_1 + r_2)^+ + \frac{1}{2}(-s_1 + s_2)^+ \\ &\leq \frac{1}{2}\sup_t (t_1 + t_2)^+ + \frac{1}{2}\sup_t (-t_1 + t_2)^+\end{aligned}$$

- $r_1 > 0, s_1 < 0$. Again r, s swap roles:

$$\begin{aligned}\frac{1}{2}(|r_1| + r_2)^+ + \frac{1}{2}(-|s_1| + s_2)^+ &= \frac{1}{2}(r_1 + r_2)^+ + \frac{1}{2}(s_1 + s_2)^+ \\ &\leq \frac{1}{2}(r_1 + r_2)^+ + \frac{1}{2}(-s_1 + s_2)^+ \\ &\leq \frac{1}{2}\sup_t (t_1 + t_2)^+ + \frac{1}{2}\sup_t (-t_1 + t_2)^+\end{aligned}$$

□

For the second ingredient (This is actually a special case of the contraction principle (Lemma 13):

Proof. (of second Lemma) Without loss of generality $l = 1$. Then we use triangle inequality:

$$\begin{aligned}\tilde{M}_n &= \frac{1}{2^{n-1}} \sum_{\varepsilon_2, \ldots, \varepsilon_n} \left\| \sum_{i=2}^n \varepsilon_i x_i \right\| \\ &\leq \frac{1}{2^{n-1}} \sum_{\varepsilon_2, \ldots, \varepsilon_n} \frac{1}{2}\left\| \varepsilon_1 x_1 + \sum_{i=2}^n \varepsilon_i x_i \right\| + \frac{1}{2}\left\| -\varepsilon_1 x_1 + \sum_{i=2}^n \varepsilon_i x_i \right\| \\ &= M_n\end{aligned}$$

□

Continuing on,

Proof. (of Proposition)

The main idea is to use triangle inequality:

$$\begin{aligned}\sum_{i=1}^{n}\delta_i|x^*(x_i)| &\le \delta\sum_{i=1}^{n}|x^*(x_i)|+\left|\sum_{i=1}^{n}(\delta_i-\delta)|x^*(x_i)|\right|\\ &\le \delta\omega_n+\left|\sum_{i=1}^{n}(\delta_i-\delta)|x^*(x_i)|\right|\end{aligned}$$

Take supremum over x^* and then expectation over δ_i and we have

$$\mathbb{E}_{\delta_i}\sup_{x^*\in B_{X^*}}\sum_{i=1}^{n}\delta_i|x^*(x_i)|\le\delta\omega_n+\mathbb{E}_{\delta_i}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}(\delta_i-\delta)|x^*(x_i)|\right|$$

Note $\delta=\frac{M_n}{\omega_n}$ so $\delta\omega_n=M_n$, and we just need to bound the second term. The approach is to introduce another set of indicators δ_i' distributed the same as δ_i but are chosen so that ε_i, δ_i, δ_i' are mutually independent.

Then note

$$\begin{aligned}\mathbb{E}_{\delta_i}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}(\delta_i-\delta)|x^*(x_i)|\right| &= \mathbb{E}_{\delta_i}\sup_{x^*\in B_{X^*}}\left|\mathbb{E}_{\delta_i'}\sum_{i=1}^{n}(\delta_i-\delta_i')|x^*(x_i)|\right|\\ \text{(Jensen)}\ &\le \mathbb{E}_{\delta_i,\delta_i'}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}(\delta_i-\delta_i')|x^*(x_i)|\right|\end{aligned}$$

where we note that the mapping $\delta_i'\mapsto\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}(\delta_i-\delta_i')|x^*(x_i)|\right|$ is convex in δ_i' (follows from convexity of $|\cdot|$ and the subadditivity of supremums:

$$\begin{aligned}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}(\delta_i-\lambda\delta_i'-(1-\lambda)\delta_i'')|x^*(x_i)|\right| &= \sup_{x^*\in B_{X^*}}\left|\lambda\sum_{i=1}^{n}(\delta_i-\delta_i')|x^*(x_i)|+(1-\lambda)\sum_{i=1}^{n}(\delta_i-\delta_i')|x^*(x_i)|\right|\\ &\le \lambda\sup_{x^*\in B_{X^*}}|...|+(1-\lambda)\sup_{x^*\in B_{X^*}}|...|\end{aligned}$$

Since $\varepsilon_i(\delta_i-\delta_i')\overset{(d)}{=}(\delta_i-\delta_i')$, we have that

$$\mathbb{E}_{\delta_i,\delta_i'}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}(\delta_i-\delta_i')|x^*(x_i)|\right|=\mathbb{E}_{\varepsilon_i,\delta_i,\delta_i'}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}\varepsilon_i(\delta_i-\delta_i')|x^*(x_i)|\right|$$

Now note that conditioning on δ_i, δ_i', since $(\delta_i-\delta_i')\in\{0,1,-1\}$, we have that by our second ingredient above (iterate it for as many x_i that become zero from $\delta_i-\delta_i'$, and note that the ε_i below makes it so that the sign distribution is still uniform on $\{\pm1\}$ for nonzero $\delta_i-\delta_i'$),

$$\mathbb{E}_{\varepsilon_i}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}\varepsilon_i(\delta_i-\delta_i')|x^*(x_i)|\right|\le 2M_n$$

and thus averaging over δ_i,δ_i' gives

$$\mathbb{E}_{\varepsilon_i,\delta_i,\delta_i'}\sup_{x^*\in B_{X^*}}\left|\sum_{i=1}^{n}\varepsilon_i(\delta_i-\delta_i')|x^*(x_i)|\right|\le 2M_n$$

Combining all the results, we have that

$$\mathbb{E}_{\delta_i} \sup_{x^* \in B_{X^*}} \left| \sum_{i=1}^{n} (\delta_i - \delta)|x^*(x_i)| \right| \leq 2M_n$$

so that

$$\mathbb{E}_{\delta_i} \sup_{x^* \in B_{X^*}} \sum_{i=1}^{n} \delta_i |x^*(x_i)| \leq \delta\omega_n + 2M_n = 3M_n$$

□

Corollary 81. *If $n M_n > 16\omega_n$ (i.e. $n\delta > 16$), we can find a subset B of $\{1, \ldots, n\}$ such that $|B| \geq \frac{n M_n}{2\omega_n}$ and*

$$\sum_{i \in B} |x^*(x_i)| \leq 4M_n \text{ for all } x^* \in B_{X^*}$$

Proof. We know there exists a choice of $\delta_i \in \{0, 1\}$ which gives the upper bound, but we want one with as many 1's as possible. We will look at bounds for the probability that the upper bound holds as well as bounds for the probability that the cardinality exceeds a certain value, and we will look for a set where both happen simultaneously.

Markov's inequality with the previous Proposition shows

$$\Pr_\delta\left\{ \sup_{x^* \in B_{X^*}} \sum_{i=1}^{n} \delta_i |x^*(x_i)| \geq 4M_n \right\} \leq \frac{3M_n}{4M_n} = \frac{3}{4}$$

and thus

$$\Pr_\delta\left\{ \sup_{x^* \in B_{X^*}} \sum_{i=1}^{n} \delta_i |x^*(x_i)| \leq 4M_n \right\} \geq \frac{1}{4}$$

Note the number of nonzero δ is precisely $\sum \delta_i$. Note that $\mathbb{E}(\sum \delta_i) = n\delta$, and

$$\mathbb{E}\left(\left| \sum \delta_i - n\delta \right|^2 \right) = \mathrm{Var}(\sum \delta_i) = \sum \mathrm{Var}(\delta_i) = n\delta(1 - \delta) \leq n\delta$$

(independence of δ_i)

Chebyshev's inequality tells us that

$$\Pr_\delta\left\{ \left| \sum \delta_i - n\delta \right| \geq \frac{n\delta}{2} \right\} \leq \frac{n\delta}{(n\delta/2)^2} = \frac{4}{n\delta}$$

and in particular,

$$\Pr_\delta\left\{ \sum \delta_i > \frac{n\delta}{2} \right\} \geq \Pr_\delta\left\{ \left| \sum \delta_i - n\delta \right| \leq \frac{n\delta}{2} \right\} \geq 1 - \frac{4}{n\delta}$$

If $n\delta > 16$, we have that the probabilities of the two events sum to > 1 and thus there must be some overlap, i.e. a choice of δ_i for which $\sum \delta_i > \frac{n\delta}{2}$ and $\sum_{i=1}^{n} \delta_i |x^*(x_i)| \leq 4M_n$ for all $x^* \in B_{X^*}$. We take B to be the index set $\{i \colon \delta_i = 1\}$, and $|B| \geq \frac{n\delta}{2} = \frac{n M_n}{2\omega_n}$

□

The Lower Bound

Now we look for a further subset that gives a lower bound in terms of the upper bound we obtained:

Proposition 82. *Suppose* $\sup\left\{\sum_{i=1}^{m}|x^*(x_i)|,\, x^*\in B_{X^*}\right\}\leq Z$. *There is a subset* A *of* $\{1,\ldots,m\}$ *such that* $|A|\geq\frac{m}{8Z}$ *and*

$$\left\|\sum_{i\in A}a_i\,x_i\right\|\geq\frac{1}{2}\sup_{i\in A}|a_i|\ \textit{for all}\ (a_i)_{i\in A}$$

Proof. For each i consider a vector x_i^* in B_{X^*} such that $x_i^*(x_i)=1$.

First, suppose A is an arbitrary subset, and let $(a_i)_{i\in A}$ be real. We note that for the index l where $|a_l|=\sup_{i\in A}|a_i|$, we have

$$\begin{aligned}\left\|\sum_{i\in A}a_i\,x_i\right\| &\geq \left|x_l^*\left(\sum_{i\in A}a_i\,x_i\right)\right|\\ &\geq |a_l||x_l^*(x_l)-\sum_{j\in A,j\neq l}|a_j||x_l^*(x_j)|\\ &= \left(\sup_{i\in A}|a_i|\right)\left(1-\sum_{j\in A,j\neq l}|x_l^*(x_j)|\right)\end{aligned}$$

Thus, if we can find a subset A for which $\sum_{j\in A,j\neq l}|x_l^*(x_j)|\leq\frac{1}{2}$ for all $l\in A$ we would have the result.

Set $\delta=\frac{1}{4Z}$ and consider an independent sequence δ_i, $i\leq m$ with $P(\delta_i=0)=1-\delta$ and $P(\delta_i=1)=\delta$. Then we note that

$$\sum_{j\leq m}|x_i^*(x_j)|\leq Z\text{ for all }i$$

and

$$\mathbb{E}\sum_{i,j\leq m,i\neq j}\delta_i\,\delta_j\,|x_i^*(x_j)|=\delta^2\sum_{i,j\leq m,i\neq j}|x_i^*(x_j)|\leq\delta^2 m\,Z=\frac{m\,\delta}{4}$$

Thus by linearity of expectation,

$$\mathbb{E}\left(\sum_{i=1}^{m}\delta_i-2\sum_{i,j\leq m,i\neq j}\delta_i\,\delta_j\,|x_i^*(x_j)|\right)\geq\frac{m\,\delta}{2}$$

This means there is a choice of δ_i for which

$$\sum_{i=1}^{m}\delta_i-2\sum_{i,j\leq m,i\neq j}\delta_i\,\delta_j\,|x_i^*(x_j)|\geq\frac{m\,\delta}{2}$$

and hence the subset $I=\{i,\,\delta_i=1\}$ where

$$|I|-2\sum_{i,j\in I,i\neq j}|x_i^*(x_j)|\geq\frac{m\delta}{2}$$

or

$$\sum_{i\in I}\left(1-2\sum_{j\in I,j\neq i}|x_i^*(x_j)|\right)\geq\frac{m\delta}{2}$$

Since each term on the left is bounded above by 1 (subtracting a positive number), we note that at least $m\delta/2$ terms are ≥ 0, and we call these indices A:

$$A=\left\{i\in I,\ \sum_{j\in I,j\neq i}|x_i^*(x_j)|\leq\frac{1}{2}\right\}$$

We have $|A|\geq\frac{m\delta}{2}=\frac{m}{8Z}$ and using this A in the beginning of the proof, we have that

$$\left\|\sum_{i\in A}a_i x_i\right\|\geq\frac{1}{2}\sup_{i\in A}|a_i|\text{ for all }(a_i)_{i\in A}$$

□

Now we just combine everything we've done:

Theorem 83. *Consider unit vectors $(x_i)_{i\leq n}$ in a Banach space, with $n\frac{M_n}{\omega_n}>16$. Then we can find a subset $A\in\{1,\ldots,n\}$ with $|A|\geq\frac{n}{64\omega_n}$ such that for any real numbers $(a_i)_{i\leq n}$ we have*

$$\frac{1}{2}\sup_{i\in A}|a_i|\leq\left\|\sum_{i\in A}a_i x_i\right\|\leq 4M_n\sup_{i\in A}|a_i|$$

Proof. If $\frac{nM_n}{\omega_n}>16$, for instance if $\frac{nM_n}{\omega_n}\geq\sqrt{\frac{n}{2}}>16$ or $n>512$ (see beginning), then Corollary 81 applies, which gives a subset of size $m\geq\frac{nM_n}{2\omega_n}$ with the upper bound $Z=4M_n$. Then Proposition 82 gives a further subset $|A|$ where the lower bound also holds, and $|A|\geq\frac{m}{8Z}\geq\frac{n}{64\omega_n}$ □

Boostrapping the upper bound

We can actually use the upper bound above to find further subsets that improve the upper bound by a square root.

Suppose for $x_1,\ldots,x_n\in X$ (linearly independent) with $\|x_i\|=1$,

$$\left\|\sum_{i=1}^n a_i x_i\right\|\leq\|a\|_\infty C$$

for all $a\in l_\infty^n$. Then there exists $y_1,\ldots,y_m\in X$, $m=\lfloor\sqrt{n}\rfloor$ linearly independent vectors with $\|y_i\|=1$ and

$$\left\|\sum_{i=1}^m a_i y_i\right\|\leq\|a\|_\infty\sqrt{C}$$

for all $a\in l_\infty^m$.

This can be proved by contradiction. Suppose for all unit $y_1, \ldots, y_m$, we can find $\|a\|_\infty \le 1$ with

$$\left\| \sum_{i=1}^{m} a_i\, y_i \right\| > \sqrt{C}$$

Then blocking x_1, ..., x_{m^2} into m groups of m vectors each, for each block $x_{km+1,\ldots,(k+1)m}$ we can find $a_{km+1,(k+1)m}$ for which

$$\left\| \sum_{i=km+1}^{(k+1)m} a_i\, x_i \right\| > \sqrt{C}$$

Let $y_k = \sum_{i=km+1}^{(k+1)m} a_i\, x_i$ for $k = 1, \ldots, m$, so that $\|y_k\| > \sqrt{C}$. Note y_k is still linearly independent, and we can find $b_1, \ldots, b_m$ such that $\|b\|_\infty \le 1$ and

$$\left\| \sum_{k=1}^{m} b_k \frac{y_k}{\|y_k\|} \right\| > \sqrt{C}$$

Then we expand out, we note

$$\begin{aligned}\sqrt{C} < \left\| \sum_{k=1}^{m} b_k \frac{y_k}{\|y_k\|} \right\| &= \frac{1}{\sqrt{C}} \left\| \sum_{k=1}^{m} \sum_{i=km+1}^{(k+1)m} \frac{\sqrt{C}}{\|y_k\|} (b_k\, a_i)\, x_i \right\| \\ &\le \frac{1}{\sqrt{C}} \left\| \sum_{i=1}^{m^2} c_i\, x_i \right\|\end{aligned}$$

where $c_i = \frac{\sqrt{C}}{\|y_k\|} b_k\, a_i$ for $km + 1 \le i \le (k+1)m$. We note $\|c\|_\infty \le 1$. This implies that $\left\| \sum_{i=1}^{m^2} c_i\, x_i \right\| > C$ which contradicts the initial assumption.

Supposing $n \frac{M_n}{\omega_n} > 16$, we had $x_1, \ldots, x_m$ with $\left\| \sum_{i=1}^{m} a_i\, x_i \right\| \le \|a\|_\infty\, 4 M_m$ and $m \ge n \frac{M_n}{2\omega_n} \ge \frac{\sqrt{n}}{2\sqrt{2}}$ By repeatedly iterating the procedure above, we can find k unit vectors $y_1, \ldots, y_k \in X$ that satisfy

$$\left\| \sum_{i=1}^{k} a_i\, y_i \right\| \le \|a\|_\infty (1 + \varepsilon/3)$$

where $k \sim m^{1/2^r} \sim n^{1/2^{r+1}}$ where r satisfies $(4 M_n)^{1/2^r} \le 1 + \varepsilon/3$ or $\frac{1}{2^r} \ln(4 M_n) \le \ln(1 + \varepsilon/3)$. This means that $k \ge C_1\, n^{C_2 \ln(1+\varepsilon)/\ln(M_n)}$.

By the triangle inequality, we can obtain the lower bound: if l is the max index where $|a_l| = \|a\|_\infty$, then

$$2|a_l| \le \left\| a_l y_l + \sum_{i \ne l} a_i\, y_i \right\| + \left\| a_l y_l - \sum_{i \ne l} a_i\, y_i \right\| \le \left\| \sum_{i=1}^{k} a_i\, y_i \right\| + |a_l|(1 + \varepsilon/3)$$

so that $\left\| \sum_{i=1}^{k} a_i\, y_i \right\| \ge \|a\|_\infty (1 - \varepsilon/3)$. Then we have shown that $y_1, \ldots, y_k$ is $(1 + \varepsilon/3)(1 - \varepsilon/3) \le 1 + \varepsilon$ isomorphic to l_∞^k, with $k \ge C_1\, n^{C_2 \ln(1+\varepsilon)/\ln(M_n)}$

Note that this says that there exists a constant $C(\varepsilon)$ such that if $\ln \ln M_n \le \delta \ln \ln n$ for some $0 < \delta < 1$, then

$$\ln \ln k \ge (1 - \delta) \ln \ln n - C(\varepsilon)$$

This is because

$$\ln k \geq \ln C_1 + C_2 \frac{\ln(1+\varepsilon)}{\ln M_n} \ln n$$

and

$$\ln \ln k \geq \ln \ln n - \ln \ln M_n + \ln(C_2 \ln(1+\varepsilon)) \geq (1-\delta) \ln \ln n - C(\varepsilon)$$

This shows

Theorem 84. *For $x_1, \ldots, x_n \in X$ unit vectors, there exists a constant $C(\varepsilon)$ such that if $\ln \ln M_n \leq \delta \ln \ln n$ for some $0 < \delta < 1$, then X contains a $(1+\varepsilon)$-isomorphic copy of l_∞^k for some k satisfying*

$$\ln \ln k \geq (1-\delta) \ln \ln n - C(\varepsilon)$$

Corollary 85. *There exists a constant $C(\varepsilon)$ such that for every n-dimensional Banach space X, either X contains an $(1+\varepsilon)$-isomorphic copy of l_2^m for some m satisfying $\ln \ln m \geq \frac{1}{2} \ln \ln n$ or X contains a $(1+\varepsilon)$-isomorphic copy of l_∞^k for some k satisfying $\ln \ln k \geq \frac{1}{2} \ln \ln n - C(\varepsilon)$.*

Proof. Take the ellipsoid of maximal volume in B_X and assume without loss of generality it is B_2^n so that $L = 1$ in Dvoretsky criteria. Then Dvoretsky criteria gives a $(1+\varepsilon)$-isomorphic copy of l_2^m for $m \geq B(\varepsilon) n M^2$ where $M = \int_{\mathbb{S}^{n-1}} \|x\| d\mu$. Note that this result is better for Banach spaces with large M.

For small M, we can show that M_n is also small, and this will give us a sizable copy of l_∞^k.

By Dvoretsky-Rogers, we can find an orthonormal basis $x_1, \ldots, x_{n/4}$ where $\|x_i\| \geq 1/2$. Then we note that

$$\begin{aligned} M = \int_{\mathbb{S}^{n-1}} \left\| \sum a_i x_i \right\| d\mu(a) &\gtrsim \int_{\mathbb{S}^{n-1}} \left\| \sum_{i=1}^{n/4} a_i \frac{x_i}{\|x_i\|} \right\| d\mu(a) \\ &= \mathbb{E}_\varepsilon \int_{\mathbb{S}^{n-1}} \left\| \sum_{i=1}^{n/4} \varepsilon_i a_i \frac{x_i}{\|x_i\|} \right\| d\mu(a) \\ &\geq \mathbb{E}_\varepsilon \left\| \sum_{i=1}^{n/4} \varepsilon_i \left(\int_{\mathbb{S}^{n-1}} a_i \, d\mu(a) \right) \frac{x_i}{\|x_i\|} \right\| \\ &\sim \mathbb{E}_\varepsilon \left\| \sum_{i=1}^{n/4} \varepsilon_i \frac{x_i}{\|x_i\|} \right\| = M_{n/4} \end{aligned}$$

If $B(\varepsilon) n M_r^2 \geq e^{\sqrt{\ln n}}$, say, then we get the copy of l_2^m with $\ln \ln m \geq \frac{1}{2} \ln \ln n$. Otherwise, we have

$$M_{n/4} \lesssim \frac{1}{n B(\varepsilon)} e^{\sqrt{\ln n}}$$

so that $\ln \ln M_{n/4} \leq \frac{1}{2} \ln \ln n$ for which the previous theorem applies, and we get a $(1+\varepsilon)$-isomorphic copy of l_∞^k with $\ln \ln k \geq \frac{1}{2} \ln \ln n - C(\varepsilon)$. $\square$