Dimension Reduction and Other Topics in Discrete Metric Geometry

William B. Johnson

Perspectives in High Dimensions, August 2010

WBJ and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space,* **Contemporary Math. 26** (1984), 189–206.

From Google:

About 9,660 for "Dvoretzky's Theorem".

About 16,500 results for "Bishop-Phelps Theorem".

About 27,900 results for "Radon-Nikodym Theorem".

About 58,200 results for "Riesz Representation Theorem".

About 45,300 results for "Fatou's Lemma".

About 75,100 results for "Johnson-Lindenstrauss Lemma".

About 194,000 results for "Schauder Fixed Point Theorem".

About 342,000 results for "Hahn-Banach Theorem".

[J-L, '84] "Given n points in Euclidean space (which we might as well take to be ℓ_2^n), what is the smallest $k = k(n)$ so that these points can be moved into k-dimensional Euclidean space via a transformation which expands or contracts all pairwise distances by a factor of at most $1+\epsilon$?"

Answer: $k(n) \leq C\ \frac{\log(n+1)}{\epsilon^2}$.

Nowadays this is called the J-L Lemma.

In fact, [J-L, '84] proved that there is a *linear* mapping T from ℓ_2^n into $\ell_2^{k(n)}$ so that for all pairs x, y from among the n points,

$$\|x-y\| \leq \|Tx-Ty\| \leq (1+\epsilon)\|x-y\|.$$

This is called the linear J-L Lemma.

There is a *linear* mapping T from ℓ_2^n into ℓ_2^k, $k = k(n) \le C\ \epsilon^{-2} \log(n+1)$, so that for all pairs x, y from among the n points,

$$\|x-y\| \le \|Tx - Ty\| \le (1+\epsilon)\|x-y\|.$$

The idea, coming from proofs of Dvoretzky's Theorem, is to use a random isometric embedding T from ℓ_2^n into ℓ_2^k. In the background is some probability space $(\Omega, \mathcal{P})$, and for each $\omega \in \Omega$, T_ω is a linear mapping from ℓ_2^n into ℓ_2^k so that for each $x \in \ell_2^n$, $\mathbb{E}\|Tx\| = \|x\|$. To almost preserve the norm of a set S of unit norm vectors in ℓ_2^n, one needs to estimate $\mathbb{P}[|\|Tx\| - \mathbb{E}\|Tx\|| > \epsilon]$ (which typically does not depend on the particular norm one vector x). If this probability is sufficiently small, a union bound argument yields that there is $\omega \in \Omega$ so that for all $x \in S$, $|\|T_\omega x\| - \mathbb{E}\|Tx\|| \le \epsilon$. For J-L, if E is a set of n points in ℓ_2^n, let S be the set of normalized differences of points in E. The random linear operator T is just a constant times a random rank k orthogonal projection on ℓ_2^n.

Conjugate any rank k orthogonal projection against the orthogonal group to get a random rank k orthogonal projection.

How did Joram and I stumble across the J-L Lemma? It was used to solve a problem from [Marcus-Pisier, '84].

Given a Banach space X, let $L(X, n)$ be the smallest constant C such that for every mapping f from an n-point subset of X into ℓ_2, there is an extension $g : X \to \ell_2$ so that

$$\text{Lip}\,(g) \le C\,\text{Lip}\,(f), \quad \text{where}$$

$$\text{Lip}\,(f) = \sup \frac{\|f(x) - f(y)\|}{d(x, y)}.$$

Kirszbraun's theorem says that $L(\ell_2, n) = 1$. Marcus and Pisier proved for $1 < p < 2$ that $L(L_p, n) \le C(p)(\log n)^{1/p-1/2}$. They asked whether always $L(X, n) \le C\,(\log n)^{1/2}$, and we showed that the J-L Lemma gives a yes answer.

The way J-L is used in compressed sensing shows that the Feichtinger conjecture is true "generically". A tight frame can be regarded as the image of an orthonormal basis under an orthogonal projection. If P is a random rank k orthogonal projection on ℓ_2^n, the expected norm of Px is about $C := \sqrt{\frac{k}{n}}$ for any unit vector x. The concentration around the expected norm guarantees that with big probability, $\frac{1}{2}C \leq \|Px\| \leq 2C$ for every $\delta(C)k$-sparse unit vector that is, for any unit vector which has at most $\delta(C)k$ non zero coordinates. This shows that $(Pe_i)_{i=1}^n$ satisfies the Feichtinger conjecture in a strong way: ANY subset of $(C^{-1}Pe_i)_{i=1}^n$ of size at most $\delta(C)k$ is 4-equivalent to an orthonormal sequence.

Say that a Banach space X satisfies the linear J-L Lemma provided there is a constant C so that for all n and all subsets E of X which contain n points there is a linear mapping T on X of rank $k \leq C \log(n+1)$ so that for all x and y in E,

$$\|x - y\| \leq \|Tx - Ty\| \leq C\|x - y\|.$$

The linear J-L Lemma is false in any L_1 space [Charikar-Sahai, '02], and in fact it is false in any L_p space, $1 \leq p \neq 2 \leq \infty$ [Lee-Mendel-Naor, '05].

[J-Naor, 2010]:
What can you say about a Banach space which satisfies the linear J-L Lemma?

Is there a Banach space which satisfies the linear J-L Lemma but is not isomorphic to a Hilbert space?

Is there a Banach space which satisfies the linear J-L Lemma but is not isomorphic to a Hilbert space?

Yes. The space $T^{(2)}$, which is the 2 convexification of Tsirelson's space [Tsirelson, '74], constructed in [Figiel-J, '74], satisfies the linear J-L Lemma.

What is easy is that a certain subspace of the 2 convexification of modified Tsirelson's space, constructed in [J, '80], satisfies the linear J-L Lemma. By later work of [Casazza-J-Tzafriri, '84] and, especially, [Casazza-Odell, '83], this space is actually equal to $T^{(2)}$.

$T^{(2)}$, like all Tsirelson-type spaces, is a Banach space of sequences whose norm is defined implicitly rather than explicitly.

Tsirelson-type spaces have played a fundamental role in constructing counterexamples in Banach space theory Gowers-Maurey; Argyros-Haydon and even for proving theorems about classical spaces, such as the existence of a distorted norm on Hilbert space [Odell-Schlumprecht, '94].

$T^{(2)}$ (defined in [Figiel-j, '74]) is the completion of the space c_{00} of sequences of scalars which have only finitely many non zero terms under the unique norm that satisfies the equation

$$\|x\|^2 = \|x\|_\infty^2 \vee \sup_{[n; n<A_1<A_2<\cdots<A_n]} \frac{1}{2}\sum_{j=1}^{n} \|A_j x\|^2.$$

$A < B$ means $\max A < \min B$ and $Ax = 1_A x$.

It is remarkable that this norm is equivalent to the 2-convexified modified Tsirelson norm defined in [J, '80], which is defined by a similar equation, but the supremum is over finite disjoint sets $n < A_j$, $1 \le j \le n$ [Casazza-Odell, '83]. Since we do not need this result, we write down the formula that defines the 2-convexified modified Tsirelson norm:

$$\|x\|^2 = \|x\|_\infty^2 \vee \sup_{[n<\cup_{j=1}^n A_j,\ A_j\cap A_i=\emptyset]} \frac{1}{2}\sum_{j=1}^{n} \|A_j x\|^2.$$

$$\|x\|^2 = \|x\|_\infty^2 \vee \sup_{[n<\cup_{j=1}^n A_j,\ A_j\cap A_i=\emptyset]} \frac{1}{2}\sum_{j=1}^n \|A_jx\|^2.$$

It is more or less obvious that if you take n disjointly supported unit vectors with supports past n, in this norm the vectors are equivalent with constant at most $\sqrt{2}$ to an orthonormal basis for ℓ_2^n. From this you get that for any n there is $r(n)$ so that any n dimensional subspace supported past $r(n)$ is 2-isomorphic to ℓ_2^n. In fact, if you change "2-isomorphic" to "$K(s)$-isomorphic", then for any iterate $\log^{[s]} n$ of $\log n$, you can get $K(s)$ to make the statement true for $r(n) = 1 \vee \log^{[s]} n$. In some stong sense, this norm is "asymptotically Hilbertian" even though the completion of c_{00} under this norm has no subspace isomorphic to ℓ_2.

$$\|x\|^2 = \|x\|_\infty^2 \vee \sup_{[n<\cup_{j=1}^n A_j,\ A_j\cap A_i=\emptyset]} \frac{1}{2}\sum_{j=1}^n \|A_j x\|^2.$$

Any 4^n-dimensional subspace supported past $\log n$ is 4-isomorphic to ℓ_2^n.

In any Banach space X, the projection constant in X of any n dimensional subspace Y is less than two times the projection constant of Y in Z for some $Y \subset Z \subset X$ with $\dim Z < 4^n$. Therefore, in $(c_{00}, \|\cdot\|)$, any n-dimensional subspace supported past $\log n$ is 4-isomorphic to ℓ_2^n and is the range of a projection whose norm is at most eight.

$$\|x\|^2 = \|x\|_\infty^2 \vee \sup_{[n<\cup_{j=1}^n A_j,\ A_j\cap A_i=\emptyset]} \frac{1}{2}\sum_{j=1}^n \|A_j x\|^2.$$

Now take any n-dimensional subspace W of c_{00} and let Y be the subspace of those $x \in W$ which are zero in the first $\log n$ coordinates. Then Y is 4-isomorphic to the Euclidean space of its dimension and is the range of a projection P of norm at most 8. $I - P$ maps W into a subspace of dimension at most $\log n$ and P maps into the 4-Euclidean space Y.

Now let E be any n-point subset of c_{00} and let W be its linear span. Apply linear J-L to get a linear operator $S : Y \to Y$ of rank at most $\log n$ s.t. for x, y in PE, $\|x-y\| \le \|Sx-Sy\| \le 5\|x-y\|$. Then the linear mapping $T := (I-P)+SP$ has rank at most $2\log n$ and distorts distances between points in E by a factor of at most 41 (well, to be safe, let's say 100 or 200).

[J-Naor, 2010]:

What can you say about a Banach space which satisfies the linear J-L Lemma?

Given a Banach space X, let $D_n(X)$ be the supremum over all n dimensional subspaces Y of X of the Banach-Mazur distances $d(Y, \ell_2^n)$.

$$d(E, F) = \inf \|T\| \cdot \|T^{-1}\|.$$

where the infimum is over all isomorphisms from E onto F.

A Banach space X is K-isomorphic to a Hilbert space iff $D_n(X) \leq K$ for all n.

For any X, $D_n(X) \leq \sqrt{n}$.

$D_n(L_p) = n^{|1/p-1/2|}$ [Lewis, '76].

$D_n(T^{(2)})$ goes to infinity slowly; slower than any iterate of $\log n$ (even slower).

$D_n(X) = \sup\{d(E, \ell_2^n) : \dim E = n;\ E \subset X\}$.

$d(Y, Z) = \inf \|T\| \cdot \|T^{-1}\|$.

[J-Naor, '10] If X satisfies linear J-L, then

$$D_n(X) \le 2^{2^{c \log^* n}},$$

where $\log^* x$ is the unique integer k such that if we define $a_1 = 1$ and $a_{i+1} = e^{a_i}$ (i.e. a_i is an exponential tower of height i), then $a_k < x \le a_{k+1}$.

More formally:

Theorem. *For every $D, K > 0$ there exists a constant $c = c(K, D) > 0$ with the following property. Let X be a Banach space such that for every $n \in \mathbb{N}$ and every $x_1, \ldots, x_n \in X$ there exists a linear subspace $F \subseteq X$, of dimension at most $K \log n$, and a linear mapping $S : X \to F$ such that $\|x_i - x_j\| \le \|Sx_i - Sx_j\| \le D\|x_i - x_j\|$ for all $i, j \in \{1, \ldots, n\}$. Then for every $n \in \mathbb{N}$ and every n-dimensional subspace $E \subseteq X$, we have*

$$d(E, \ell_2^n) \le 2^{2^{c \log^*(n)}}. \tag{1}$$

$D_n(X) = \sup\{d(E, \ell_2^n) : \dim E = n;\ E \subset X\}.$

$d(Y, Z) = \inf \|T\| \cdot \|T^{-1}\|.$

[J-Naor, '10] If X satisfies linear J-L, then

$$D_n(X) \le 2^{2^{c \log^* n}},$$

where $\log^* x$ is the unique integer k such that if we define $a_1 = 1$ and $a_{i+1} = e^{a_i}$ (i.e. a_i is an exponential tower of height i), then $a_k < x \le a_{k+1}$.

The somewhat technical proof uses harmonic analysis on $\{-1, 1\}^n$ and standard results from the local theory of Banach spaces. Part of it is based on the ideas in [Charikar-Sahai, '02], [Lee-Mendel-Naor, '05]. Whether it is the "right" result is open. For $T^{(2)}$ it follows easily from [Bellenot, '84] that

$$D_n(T^{(2)}) \ge 2^{c\alpha(n)},$$

where $\alpha(n) \to \infty$ is the inverse Ackermann function (which is much smaller than $2^{2^{c \log^*(n)}}$).

Positive results on dimension reduction for spaces other than Hilbert spaces are few. [Matousek, '96] proved that if $1 > a > 0$ and E is any n-point metric space, then E embeds into ℓ_∞^k with distortion at most C/a for some $k \leq C\, n^a$, and up to the constant C this is best possible.

Arguably the most important space other than ℓ_2 for having results on dimension reduction is L_1. Until this year, the only positive result was that n points in L_1 embed into $\ell_1^{n \log n}$ with constant distortion [Schechtman, '87]. But, following up on the idea introduced there, in [Bourgain-Lindenstrauss-Milman, '89], [Talagrand, '90] it was proved that every n-dimensional subspace of L_1 linearly embeds into $\ell_1^{n \log n}$ with constant distortion. That is, we could not do better for n-points than what we can do for their linear span!

But now we know that n points in L_1 must $1 + \epsilon$-embed into ℓ_1^m with $m \leq Cn/\epsilon^2$ [Newman-Rabinovich, '10].

[Brinkman-Charikar, '05] made a breakthrough on getting a lower bound for dimension reduction in L_1 which goes far beyond showing that the J-L Lemma is false for L_1. The precise statement of their theorem is that for each n there are subsets A_n of ℓ_1 of cardinality $|A_n| = n$ so that if $\alpha > 0$ and $f_n : A_n \to \ell_1^{\lceil n^\alpha \rceil}$, then $\text{Lip}\,(f_n)\text{Lip}\,(f_n^{-1}) \geq c\alpha^{-1/2}$ for some universal $c > 0$. A much simpler proof of this was given in [Lee-Naor, '04] and some further simplifications were made in [J-Schechtman, '10]. I'll sketch the proof from [J-Schechtman, '10].

The relevant subsets of L_1 for getting the lower bound on dimension reduction are the *diamond graphs* D_n with the graph metric, which all embed into L_1 with distortion 2. D_0 has two vertices joined by one edge. D_{n+1} is obtained from D_n by erasing each edge $[u, v]$ in D_n, adding two new points x, y for each edge $[u, v]$, and adding edges $[u, x]$, $[x, v]$, $[u, y]$, and $[y, v]$. Thus D_1 is a square, D_2 begins to look like a diamond, and D_9 really sparkles.

Non embedability of D_n into ℓ_1^k

A frequently used technique in Banach space theory (and used by [Lee-Naor, '04]) is to replace ℓ_1^k by ℓ_p^k where $p' = \log k$ and where $1/p + 1/p' = 1$. This is no loss because for this value of p,

$$d(\ell_1^k, \ell_p^k) = d(\ell_\infty^k, \ell_{p'}^k) \le k^{1/p'} = k^{1/\log k} = e.$$

The gain from this is that L_p is uniformly convex for $1 < p < \infty$, and its modulus of uniform convexity is known. In particular, L_p does not contain the "graph square" D_2 isometrically, and in fact you can easily estimate the the distortion needed in order to embed D_2 into L_p.

The modulus of uniform convexity of X is the function $\delta = \delta_X : (0,2) \to [0,1]$ defined by

$$\delta(\varepsilon) = \inf\{1 - \|\frac{x+y}{2}\| \; ; \; \|x\|, \|y\| \le 1, \; \|x-y\| \ge \varepsilon\}.$$

$\delta(\varepsilon) = \inf\{1 - \|\frac{x+y}{2}\| \,;\ \|x\|, \|y\| \leq 1,\ \|x-y\| \geq \varepsilon\}$.

$D_1 = \{0,1\}^2$ with edges $[(0,0),(0,1)]$, $[(0,0),(1,0)], [(0,1),(1,1)], [(1,0),(1,1)]$.

Lemma 1 *Let X be a normed space and $f : D_1 \to X$ with* $\mathrm{Lip}\,(f^{-1}) \leq 1$ *and* $\mathrm{Lip}\,(f) \leq M$. *Then* $\|f(1,1) - f(0,0)\| \leq 2M(1 - \delta(\frac{2}{M}))$.

Proof: Without loss of generality we may assume $f(0,0) = 0$. Denote $x = f(1,1)$ and $x_1 = x - f(1,0)$, $x_2 = f(1,0)$, $x_3 = x - f(0,1)$, $x_4 = f(0,1)$. Then, $1 \leq \|x_i\| \leq M$ for $i = 1,2,3,4$. Since $\|\frac{x_2}{M} - \frac{x_4}{M}\| \geq \frac{2}{M}$, we get that

$$1 - \frac{\|x_2 + x_4\|}{2M} \geq \delta(\frac{2}{M}).$$

Similarly,

$$1 - \frac{\|x_1 + x_3\|}{2M} \geq \delta(\frac{2}{M}).$$

Consequently,

$$2(1 - \delta(\frac{2}{M})) \geq \frac{\|x_1 + x_2 + x_3 + x_4\|}{2M} = \frac{\|x\|}{M}.$$

$\delta(\varepsilon) = \inf\{1 - \|\frac{x+y}{2}\| \, ; \; \|x\|, \|y\| \le 1, \; \|x - y\| \ge \varepsilon\}$.

Lemma 1:

$f : D_1 \to X$, $\mathrm{Lip}\,(f^{-1}) \le 1$, and $\mathrm{Lip}\,(f) \le M$.
Then $\|f(1,1) - f(0,0)\| \le 2M(1 - \delta(\frac{2}{M}))$.

Applying the lemma we get that if M_n is the best constant M such that there is an embedding f of D_n into X with $d_{D_n}(x,y) \le \|f(x) - f(y)\| \le M d_{D_n}(x,y)$, then $M_{n-1} \le M_n(1 - \delta_X(\frac{2}{M_n}))$. Indeed, by the Lemma the distance between the images of the top and the bottom vertices of any sub diamond of level $n-1$ is at most $2M_n(1 - \delta_X(\frac{2}{M_n}))$ (and at least 2). The collection of all top and bottom vertices of level $n-1$ of D_n is isometric to D_{n-1}: The distance between any two points of that subset of D_n is exactly twice the distance of the vertices of D_{n-1} they developed from. This implies that $M_{n-1} \le M_n(1 - \delta_X(\frac{2}{M_n}))$ or

$$M_n - M_{n-1} \ge M_{n-1}\delta_X(\frac{2}{M_n}).$$

$$\delta_X(\varepsilon) = \inf\{1 - \|\frac{x+y}{2}\|;\ \|x\|, \|y\| \leq 1,\ \|x-y\| \geq \varepsilon\}.$$

M_n: best constant M s.t. $\exists f : D_n \to X$ with $d_{D_n}(x,y) \leq \|f(x) - f(y)\| \leq M d_{D_n}(x,y)$. Then

$$M_n - M_{n-1} \geq M_{n-1} \delta_X(\frac{2}{M_n}).$$

So if $\delta_X(\varepsilon) \geq c\varepsilon^p$ then

$$M_n \geq c2^p \sum_{k=1}^{n} M_k^{-p+1} + M_0 \geq c2^p n M_n^{-p+1},$$

implying $M_n \geq 2c^{1/p} n^{1/p}$.

$$\delta_X(\varepsilon) = \inf\{1 - \|\frac{x+y}{2}\|;\ \|x\|, \|y\| \le 1,\ \|x-y\| \ge \varepsilon\}.$$

M_n: best M s.t. $\exists f : D_n \to X$ with $d_{D_n}(x,y) \le \|f(x) - f(y)\| \le M d_{D_n}(x,y)$. If $\delta_X(\varepsilon) \ge c\varepsilon^p$ then

$$M_n \ge 2c^{1/p} n^{1/p}.$$

Corollary 1 *[Brinkman-Charikar, '05] If D_n Lipschitz embeds into ℓ_1^k with distortion K then $k \ge |D_n|^{\beta/K^2}$, for a universal $\beta > 0$.*

Set $p' = \frac{p}{p-1} = \log k$ (i.e., $p = 1 + \frac{1}{\log k}$), so that $d(\ell_1^k, \ell_p^k) \le e$. Classical fact: for $1 < p \le 2$,

$$\delta_{L_p}(\epsilon) \ge \alpha(p-1)\epsilon^2,$$

for a universal $\alpha > 0$. Thus

$$eK \ge 2\alpha^{1/2}(p-1)^{1/2} n^{1/2}.$$

Plugging in $p = 1 + \frac{1}{\log k}$ we get $\log k \ge \beta n/K^2$ for a universal $\beta > 0$. This gives the corollary because

$$|D_n| = 2 + 2\frac{4^n - 1}{3}.$$

Another use for the diamond graphs

Let T_n be the dyadic tree of depth n. [Bourgain, '86] proved that a Banach space X is not isomorphic to a uniformly convex space iff T_n embeds into X with distortion independent of n.

[J-Schechtman, '10] proved that a Banach space X is not isomorphic to a uniformly convex space iff D_n embeds into X with distortion independent of n iff the Laakso graph L_n embeds into X with distortion independent of n. L_n "looks" a lot like D_n but has bounded geometry.

It is not clear how to derive either result from the other.