#2 Low-distortion embedding of finite metric spaces into ℓ_1, ℓ_2, and ℓ_∞

A mapping $\varphi : X \to X'$, where (X, d) and (X', d') are metric spaces, has distortion at most β, or is called a β-embedding, where $\beta \geq 1$, if there is an $c \in (0, +\infty)$ such that for all $x, y \in X$

$$c\, d(x,y) \leq d'(\varphi(x), \varphi(y)) \leq c\beta\, d(x,y).$$

If X' is a normed space, we usually require $c = \frac{1}{\beta}$ or $c = 1$.

If $c = 1$, we obtain

$$d(x,y) \leq d'(\varphi(x), \varphi(y)) \leq \beta\, d(x,y). \quad \text{(dilation)}$$

If $c = \frac{1}{\beta}$, we obtain

$$\frac{1}{\beta}\, d(x,y) \leq d'(\varphi(x), \varphi(y)) \leq d(x,y) \quad \text{(contraction)}$$

Mappings with bounded distortion are also called bi-Lipschitz mappings.

(i) Bourgain's, Matousek's, and Rao's low-distortion embeddings theorems;

(ii) The Johnson-Lindenstrauss flattening lemma;

(iii) Probabilistic embeddings into tree-metrics;

(iv) Applications of embeddings;

(v) Proof of Bourgain's theorem.

Theorem 15 (Bourgain, 1985) Any n-point metric space (V_n, d) can be embedded into ℓ_1^m (in fact, into every ℓ_p) with distortion $O(\log n)$ and dimension $m = O(\log^2 n)$

Idea of the proof (more details below):

Set $\ell = \lfloor \log_2 n \rfloor$ and $q = \lfloor C \log n \rfloor$ (C is a suitable constant)

Consider an embedding into $\ell_1^{\ell q}$ with coordinates indexed by $i = 1, \dots, \ell$ and $j = 1, \dots, q$:

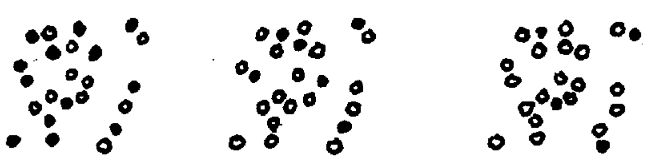
for each i, j, select a subset $A_{ij} \subseteq V_n$ by putting each $x \in V_n$ into A_{ij} with probability $\frac{1}{2^i}$, all random choices being mutually independent.

Set $\varphi(x)_{ij} = d(x, A_{ij}) := \min\{d(x,a) : a \in A_{ij}\}$

and

$$\varphi(x) = (\varphi(x)_{11}, \dots, \varphi(x)_{1q}, \dots, \varphi(x)_{\ell 1}, \dots, \varphi(x)_{\ell q})$$

Then $\varphi : V_n \longrightarrow \ell_1^{\ell q}$ is an $O(\log n)$-distortion ℓ_1-embedding with propability at least $1/2$.



A_{*1} $\qquad$ A_{*2} $\qquad$ A_{*3}

Remark: The dimension m of ℓ_1^m in the original Bourgain' proof was exponential. It has been reduced to $O(\log^2 n)$ by Linial, London, and Rabinovich (1995) using Chernoff bounds.

Theorem 16 (Matoušek, 1996) For an integer $b>0$ set $\beta = 2b-1$. Then any n-point metric space (V_n, d) can be embedded into ℓ_∞^m with distortion β, where $m = O(b n^{\frac{1}{b}} \log n)$

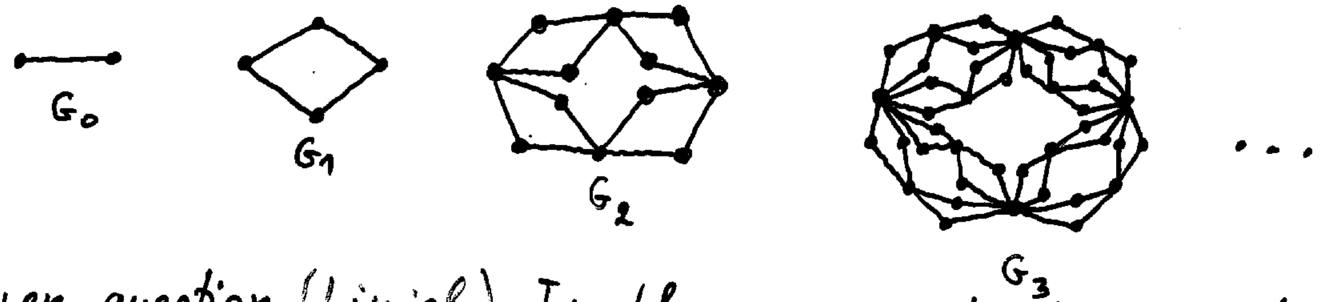
Remark: Recall that (V_n, d) embeds into ℓ_∞^n without any distortion according to Frechet's theorem.

Remark: For the special case $\beta = O(\log n)$, Matousek's result implies that (V_n, d) embeds into $\ell_\infty^{O(\log^2 n)}$ with $O(\log n)$ distortion.

Planar metric: A finite metric space (V_n, d) is planar if (V_n, d) isometrically embeds (without any distortion!) into a planar graph.

Theorem 17 (Rao, 1999) Any n-point planar metric can be embedded into ℓ_2 with distortion $O(\sqrt{\log n})$.

Remark: Newman & Rabinovich (2002) established that the $O(\sqrt{\log n})$ bound is sharp for the diamond graph (known also as Laakso's fractal)



Open question (Linial) Is there a constant C such that any planar metric embeds into ℓ_1 with distortion $\leq C$?

Open question (Linial, Rabinovich) Characterize planar metrics. In particular, given a metric, can one decide in polynomial time whether it is a planar metric?

Doubling dimension of a metric space (Assouad, 1983)

The doubling constant of a metric space (X,d) is the smallest value λ such that every ball in X can be covered by λ balls of half the radius.

The doubling dimension of (X,d) is $\log_2 \lambda$.

The doubling dimension of m-dimensional ℓ_p space is roughly m.

Theorem 18 (Assouad, 1983) If the doubling dimension of a metric space (X,d) is bounded, then for any $0<\alpha<1$ the metric space (X, d^α) (the snowflaked version of d) embeds into ℓ_2^m with distortion β, where m and β depend only on the doubling dimension of X.

Remark: Assouad (1983) conjectured that Theorem 18 holds even when $\alpha=1$, but Semmes (1996) disproved this conjecture. Gupta, Krauthgamer, Lee (2003) established that the result holds for trees with doubling dimension $< \infty$.

Edit (or Levenstein) distance: Σ – a finite alphabet,
Σ^* all strings with symbols from Σ.
For $s, s' \in \Sigma^*$, $d_E(s,s')$ - the minimum number of edit operations (insertion, deletion, substitution) transforming s into s'.

Open question (Indyk) Is there a constant c such that the metric space (Σ^*, d_E) embeds into ℓ_1 with distortion $\leq c$?

Theorem 19 (Johnson-Lindenstrauss flattening lemma, 1984)
Let X be an n-point set in a Euclidean space, and let $\varepsilon \in (0,1]$ be given. Then X can be embedded into $\ell_2^{O(\varepsilon^{-2} \log n)}$ with distortion $(1+\varepsilon)$.

Remark: Theorem 19 can be viewed as a dimensionality reduction result: a set of points in a high-dimensional space is mapped to a space with low dimension, while (approximately) preserving important characteristics of the pointset.

Idea of the proof: Set $m := \frac{200 \ln n}{\varepsilon^2}$ and assume $m<n$ (otherwise, there is nothing to prove).

Let L be a random m-dimensional linear subspace of ℓ_2^n.

Let $p : \ell_2^n \to L$ be the orthogonal projection onto L.

Claim: For any two distinct points $x, y \in \ell_2^n$, the condition

$$(1-\tfrac{\varepsilon}{3})\mu \|x-y\|_2 \le \|p(x)-p(y)\|_2 \le (1+\tfrac{\varepsilon}{3})\mu \|x-y\|_2 \qquad (*)$$

is violated with probability at most n^{-2}.

Since $|X|=n$ and $\frac{n(n-1)}{2} < n^2$ pairs of distinct $x, y \in X$, there exists some L such that $(*)$ holds for all $x, y \in X$.
In this case, the mapping $p : X \to L$ has distortion $\beta \le \frac{1+\varepsilon/3}{1-\varepsilon/3} < 1+\varepsilon$ for $\varepsilon \le 1$.

The value of μ in previous claim is defined by:

<u>Lemma</u> (concentration of the length of the projection)

For a unit vector $x \in S^{n-1}$, let $f(x) = \sqrt{x_1^2 + \ldots + x_m^2}$ be the length of the projection of x on the subspace L_0 spanned by the first m coordinates. Then $f(x)$ is sharply concentrated around a suitable number $\mu = \mu(n,m)$

$$\Pr[f(x) \geq \mu + t] \leq 2e^{-t^2 n/2} \text{ and } \Pr[f(x) \leq \mu - t] \leq 2e^{-t^2 n/2},$$

where $\Pr$ is the uniform probability measure of the sphere S^n

<u>Theorem 20</u> (Brinkman, Charikar, 2003; see Lee & Naor, 2004 for a very short proof) There exists an n-point subset $X \subseteq \ell_1$ such that for any $\beta > 1$, if X embeds into ℓ_1^m with distortion β, then $m \geq n^{\Omega(1/\beta^2)}$. In other words, the dimensionality reduction is impossible in ℓ_1-metrics.

<u>Idea of the proof of Lee & Naor</u>:

Let G_k be the kth diamond graph with all edges of length 2^{-k}.

(i) G_k can be embedded with constant distortion into ℓ_1.

(ii) using simple counting and the "short diagonal lemma" it is shown that for every $1 < p \leq 2$, any embedding of G_k into ℓ_p incurs distortion $\geq \sqrt{1 + (p-1)k}$

(iii) ℓ_1^m is $O(1)$-isomorphic to ℓ_p^m, when $p = 1 + \frac{1}{\log m}$

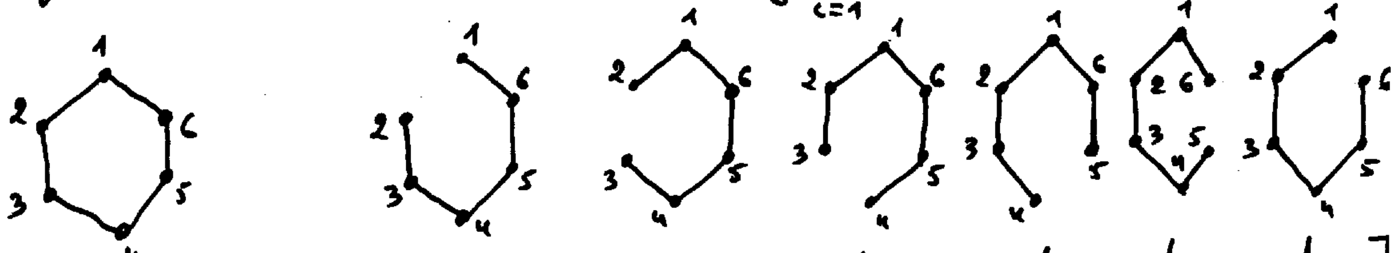
Dominating metric: let d, d' be metrics on the same set X. Then d' dominates d if $d'(x,y) \geq d(x,y)$ for all $x,y \in X$

Let $\mathcal{D}$ be a finite family of metrics on the same set X and let $\mathcal{P}$ be a probability distribution over $\mathcal{D}$. Then $(\mathcal{D}, \mathcal{P})$ β-probabilistically approximates a metric d on X if

(i) every metric $d_i \in \mathcal{D}$ dominates d;

(ii) $\forall x,y \in X$, $E_{d_i \in \mathcal{D}}[d_i(x,y)] = \sum_{i=1}^{k} p_i d_i(x,y) \leq \beta d(x,y)$.

Example: Let d be the graph metric of the cycle C_n. Let d_i be the graph metric of the path obtained by removing the i-th edge of C_n. The $\frac{1}{6}\sum_{i=1}^{6} d_i \leq 3d$



$$3 \times d \geq \frac{1}{6}\left[d_1 + d_2 + d_3 + d_4 + d_5 + d_6 \right]$$

Theorem 21 (Fakcharoenphol, Rao, Talwar, 2003) Let (X,d) be a finite metric space. Then there exists a set $\mathcal{D} = \{d_1, \dots, d_k\}$ of tree metrics on X and a probability distribution $\mathcal{P}$ over $\mathcal{D}$ such that $(\mathcal{D}, \mathcal{P})$ $O(\log n)$-probabilistically approximates d, where $n = |X|$. If d is a metric of a weighted graph G, then the tree metrics can be chosen to be spanning trees of G.

Remark: Theorem 21 improves on previous result of Bartal (1998).

<u>Theorem 15</u> (Bourgain, 1985) $(V_n, d) \xhookrightarrow{O(\log n)} \ell_1^{O(\log^2 n)}$

<u>Proof</u>: First consider the following one-dimensional embedding of (V_n, d): pick $S \subseteq V_n$ and $\forall v \in V_n$ set

$$\sigma(v) = \min\{d(v,s) : s \in S\}.$$

<u>Lemma 1</u>: $|\sigma(u) - \sigma(v)| \leq d(u,v) \quad \forall u,v \in V_n$

<u>Proof</u>: Let s_1 and s_2 be the closest vertices of S to u and v, resp. Assume w.l.o.g. that $d(s_1,u) \leq d(s_2,v)$. Then

$$|\sigma(u) - \sigma(v)| = d(s_2,v) - d(s_1,u) \leq d(s_1,v) - d(s_1,u) \leq d(u,v),$$

the last inequality follows from triangle inequality □

Now, pick ℓ subset of V_n, $S_1, \dots, S_\ell$, and <u>define the ith coordinate</u> of $v \in V_n$ to be $\sigma_i(v) = \min_{s \in S_i} d(s,v)/\ell$. Let $\ell = \log_2 n + 1$; for each $2 \leq i \leq \ell$, <u>set S_i is formed by picking each vertex of V_n with probability $1/2^i$</u>.

From Lemma 1 we conclude that

$$\|\sigma(u) - \sigma(v)\|_1 = \sum_{i=1}^{\ell} |\sigma_i(u) - \sigma_i(v)| \leq d(u,v).$$

Now we will ensure that a single distance $d(u,v)$ is not overshrunk. For this, we consider the expected contribution of set S_i $E[|\sigma_i(u) - \sigma_i(v)|]$ to the ℓ_1-distance between u and v.

Let $B(x,r) = \{v \in V : d(x,v) \leq r\}$ denote the ball of radius r around x.

Lemma 2: If for some choice of $r_1 \geq r_2 \geq 0$ and constant c,

$$\Pr[(S_i \cap B(u, r_1) = \emptyset) \text{ and } (S_i \cap B(v, r_2) \neq \emptyset)] \geq c,$$

then the expected contribution of S_i is $\geq c(r_1 - r_2)/\ell$.

Proof: Under the event described, $d(u, S_i) \geq r_1$ and $d(v, S_i) \leq r_2$. Then $\sigma_i(u) \geq r_1/\ell$ and $\sigma_i(v) \leq r_2/\ell$. Therefore,

$$|\sigma_i(u) - \sigma_i(v)| \geq (r_1 - r_2)/\ell,$$

thus the expected contribution of S_i is $\geq c(r_1 - r_2)/\ell$. $\square$

For each set S_i we will define r_1 and r_2 such that the statement of Lemma 2 holds.

Lemma 3: For $1 \leq t \leq \ell - 1$, let A and B be disjoint subsets of V_n such that $|A| < 2^t$ and $|B| \geq 2^{t-1}$. Form set S by picking each vertex of V_n independently with probability $p = 1/2^{t+1}$. Then,

$$\Pr[(S \cap A = \emptyset) \text{ and } (S \cap B \neq \emptyset)] \geq \frac{1}{2}(1 - e^{-\frac{1}{4}})$$

Proof: $\Pr[S \cap A = \emptyset] = (1-p)^{|A|} \geq (1 - p|A|) \geq \frac{1}{2}$

$\Pr[S \cap B = \emptyset] = (1-p)^{|B|} \leq e^{-p|B|} \leq e^{-\frac{1}{4}}$ (we used $1 - x \leq e^{-x}$)

whence

$$\Pr[S \cap B \neq \emptyset] = 1 - (1-p)^{|B|} \geq 1 - e^{-\frac{1}{4}}.$$

Since $A \cap B = \emptyset$, the events $[S \cap A = \emptyset]$ and $[S \cap B \neq \emptyset]$ are independent, thus the desired probability is $\geq \frac{1}{2}(1 - e^{-\frac{1}{4}})$. $\square$

Set $c = \frac{1}{2}(1 - e^{-\frac{1}{4}})$.

For $0 \le t \le \ell - 1 = [\log_2 n]$, define

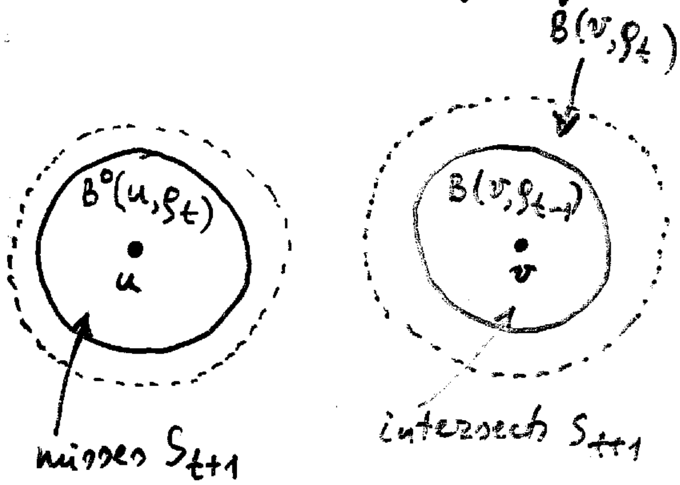
$$\rho_t = \min\{\rho \ge 0 : |B(u,\rho)| \ge 2^t \text{ and } |B(v,\rho)| \ge 2^t\}.$$

Let $\hat{t} = \max\{t : \rho_t < d(u,v)/2\}$; clearly $\hat{t} \le \ell - 2$.

Let $B^\circ(x,r) = \{s \in V : d(x,s) < r\}$ – the open ball.

__Lemma 4__: For $1 \le t \le \hat{t}$, the expected contribution of S_{t+1} is at ~~most~~ least $c \cdot \frac{\rho_t - \rho_{t-1}}{\ell}$. For $t = \hat{t}+1$, the expected contribution of S_{t+1} is at most $\frac{c}{\ell}\left(\frac{d(u,v)}{2} - \rho_{t-1}\right)$.

__Proof__: We will prove only the first assertion, i.e. $1 \le t \le \hat{t}$. By definition of ρ_t, at least one of the open balls $B^\circ(u,\rho_t)$, $B^\circ(v,\rho_t)$ contains fewer than 2^t vertices. Assume w.l.o.g. $|B^\circ(u,\rho_t)| < 2^t$. By definition, $|B(v,\rho_{t-1})| \ge 2^{t-1}$. Since $\rho_{t-1} < \rho_t < d(u,v)/2$, the two sets $B^\circ(u,\rho_t)$ and $B(v,\rho_{t-1})$ are disjoint. By Lemma 3, the probability that S_{t+1} is disjoint from $B^\circ(u,\rho_t)$ and intersects $B(v,\rho_{t-1})$ is at least c. Since $B^\circ(u,\rho_t)$ is a ball centered at u and radius $< \rho_t$ the assertion follows from Lemma 2.1



__Lemma 5__: The expected contribution of all sets $S_2, \ldots, S_\ell$ is at ~~most~~ least $\frac{c}{2\ell} d(u,v)$.

__Proof__: By Lemma 4, the expected contribution of all sets $S_2, \ldots S_\ell$ is at least the following telescoping sum:

$$\frac{c}{\ell}\left[(\rho_1 - \rho_0) + (\rho_2 - \rho_1) + \ldots + \left(\frac{d(u,v)}{2} - \rho_{\hat{t}}\right)\right] = \frac{c}{2\ell} d(u,v). \quad \square$$

Lemma 6: $\Pr\left[\text{contribution of all sets is} \geq \frac{c\,d(u,v)}{2\ell}\right] \geq \frac{c/2}{1-c/2}$

Proof: follows from Lemma 5.

Chernoff bound: let $X_1, \ldots, X_N$ be independent Bernoulli trials with $\Pr[X_i = 1] = p$ and let $X = \sum_{i=1}^{N} X_i$ ($E[X] = Np$). Then for $0 < \varepsilon \leq 1$,

$$\Pr[X < (1-\varepsilon)Np] < e^{-\frac{\varepsilon^2 Np}{2}}.$$

Pick sets $S_2, \ldots, S_\ell$ using probabilities specified above, independently $N = O(\log n)$ times each. Call the sets so obtained S_{ij}, $1 \leq i \leq \ell$, $1 \leq j \leq N$. Consider the $\ell \cdot N = O(\log^2 n)$ dimensional embedding of (V_n, d) with respect to these $\ell \cdot N$ sets.

Lemma 7: $\Pr\left[\|\sigma(u) - \sigma(v)\|_1 \geq \frac{p\,c\,d(u,v)}{4\ell}\right] \geq 1 - \frac{1}{2n^2}$ $\left(p = \frac{c}{2-c}\right.$

Proof: Think of picking sets $S_2, \ldots, S_\ell$ once as a single Bernoulli trial (thus we have N such trials). A trial succeed if the contribution of all sets is $\geq \frac{c\,d(u,v)}{2\ell}$; the probability of success is $\geq p = \frac{c}{2-c}$ by Lemma 6.

Using Chernoff bound with $\varepsilon = \frac{1}{2}$, the probability that at most $Np/2$ of these trials succeed is $\leq e^{Np/8} \leq \frac{1}{2n^2}$ for $N = O(\log n)$. If at least $Np/2$ trials succeeds the ℓ_1-distance between $\sigma(u)$ and $\sigma(v)$ will be $\geq \frac{p\,c\,d(u,v)}{4\ell} = \frac{d(u,v)}{O(\log n)}$.

Adding the error prob. for all $\frac{n(n-1)}{2}$ pairs:

Theorem: With probability $\geq \frac{1}{2}$ this $O(\log^2 n)$ dimensional embedding has distortion $O(\log n)$