Measured Descent: A new embedding method for finite metrics

R. Krauthgamer, J. Lee, M. Mendel, A. Naor (2004)

Metric Embeddings Seminar

March 6, 2007

Presentation by Evan Ettinger

The Main Result of KLMN04

Theorem (Padded embedding theorem)

Every n-point metric space (X, d) with doubling constant λ_x can be embedded in L_p for $1 \leq p < \infty$ with distortion:

$$\mathcal{O}((\log \lambda_x)^{1-1/p}(\log n)^{1/p})$$

- $p = 2 : \mathcal{O}(\sqrt{\log \lambda_x \log n})$
  - Better than Bourgain when $\log \lambda_x \leq \mathcal{O}(\log n)$
- Examples exist for tightness of bound when:
  - $\lambda_x = \mathcal{O}(1)$ – Laakso Graph
  - $\lambda_x = n^{\Omega(1)}$ – Expanders
  - Open problem: Tight for $\lambda_x \in \{c_1, \ldots, n^{c_2}\}$ where $c_1 \in \mathbb{N}, 0 < c_2 < 1$?

The Main Result of KLMN04

Theorem (Padded embedding theorem)

Every n-point metric space (X, d) with doubling constant λ_x can be embedded in L_p for $1 \leq p < \infty$ with distortion:

$$\mathcal{O}((\log \lambda_x)^{1-1/p}(\log n)^{1/p})$$

- $p = 2 : \mathcal{O}(\sqrt{\log \lambda_x \log n})$
  - Better than Bourgain when $\log \lambda_x \leq \mathcal{O}(\log n)$
- Examples exist for tightness of bound when:
  - $\lambda_x = \mathcal{O}(1)$ – Laakso Graph
  - $\lambda_x = n^{\Omega(1)}$ – Expanders
  - Open problem: Tight for $\lambda_x \in \{c_1, \ldots, n^{c_2}\}$ where $c_1 \in \mathbb{N}, 0 < c_2 < 1$?

The Main Result of KLMN04

Theorem (Padded embedding theorem)

Every n-point metric space (X, d) with doubling constant λ_x can be embedded in L_p for $1 \leq p < \infty$ with distortion:

$$\mathcal{O}((\log \lambda_x)^{1-1/p}(\log n)^{1/p})$$

- $p = 2 : \mathcal{O}(\sqrt{\log \lambda_x \log n})$
  - Better than Bourgain when $\log \lambda_x \leq \mathcal{O}(\log n)$
- Examples exist for tightness of bound when:
  - $\lambda_x = \mathcal{O}(1)$ – Laakso Graph
  - $\lambda_x = n^{\Omega(1)}$ – Expanders
  - Open problem: Tight for $\lambda_x \in \{c_1, \ldots, n^{c_2}\}$ where $c_1 \in \mathbb{N}, 0 < c_2 < 1$?

The Main Result of KLMN04

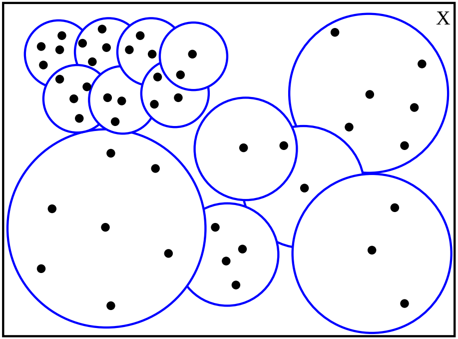
Theorem (Padded embedding theorem)

Every n-point metric space (X, d) with doubling constant λ_x can be embedded in L_p for $1 \leq p < \infty$ with distortion:

$$\mathcal{O}((\log \lambda_x)^{1-1/p}(\log n)^{1/p})$$

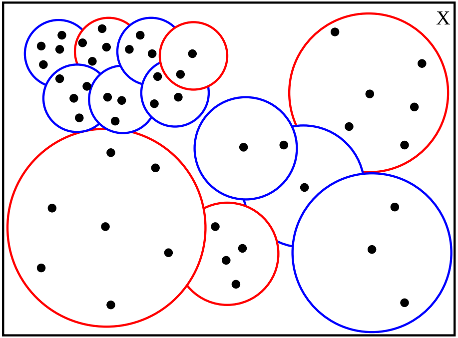
- $p = 2 : \mathcal{O}(\sqrt{\log \lambda_x \log n})$
  - Better than Bourgain when $\log \lambda_x \leq \mathcal{O}(\log n)$
- Examples exist for tightness of bound when:
  - $\lambda_x = \mathcal{O}(1)$ – Laakso Graph
  - $\lambda_x = n^{\Omega(1)}$ – Expanders
  - Open problem: Tight for $\lambda_x \in \{c_1, \ldots, n^{c_2}\}$ where $c_1 \in \mathbb{N}, 0 < c_2 < 1$?

Main Idea of Embedding



- Partition X at varying "speeds" and create a Bourgain-style embedding.
- Speed depends on density of points in a particular region of space

Main Idea of Embedding



- Each coordinate will be distance to a set W_t that is composed of many different scales "glued" together.

Padded Decompositions

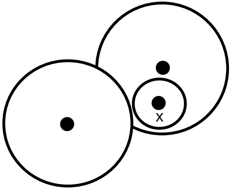
- One of the main ideas we will use in the embedding is the following:

Definition

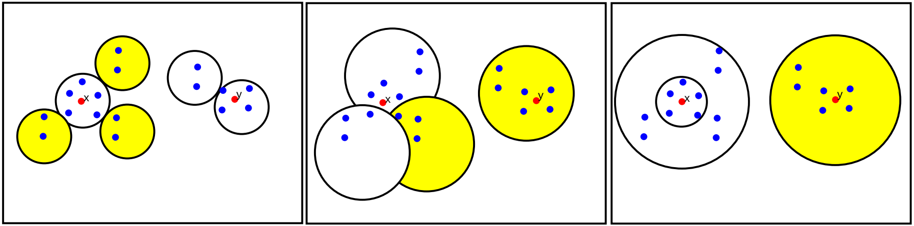
Let $\varepsilon : X \times \mathbb{Z} \to (0,1]$ and $\Delta \in \{2^u : u \in \mathbb{Z}\}$. An *$\varepsilon$-padded decomposition bundle* is a series of partitions P_u of X where each P_u satisfies:

1. For all blocks $B \in P_u$, $diam(B) \leq \Delta$.
2. $\forall x \in X$, $Pr[B(x, \varepsilon(x,\Delta)\Delta) \not\subseteq P_u(x)] \leq \frac{1}{2}$.

- Small ball around x doesn't escape partition block.

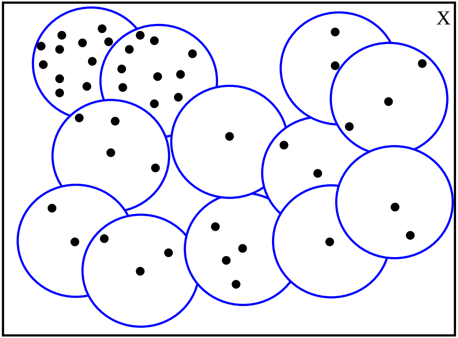


Why Padded Decompositions?



- Consider the scale $\Delta \approx \frac{1}{2}d(x, y)$
- Pr[Pick y's block]Pr[Don't pick x's block]Pr[x is padded] $> \alpha$.
- In expectation, we preserve a fraction of $d(x, y)$.

Partition Algorithm



$P_u = \mathit{Partition}(\Delta = 2^u, X)$:

1. Pick u.a.r $R \in [\Delta/2, \Delta]$.
2. Pick permutation π of X u.a.r.
3. For $i = 1, \ldots, n$
    1. Proceed according to π and assign all elements in $B(\pi(i), R)$ that were not previously assigned to the same block.

- P_u is the partition at scale $\Delta = 2^u$ and $P_u(x)$ is the *block* in which x was placed.

Padded Decomposition Lemma

Lemma (Padded Decomposition Lemma)

Let X be any metric space, and let μ be any non-degenerate measure on X. Then there exists a padded decomposition bundle on X where

$$\varepsilon(x, \Delta) = \frac{1}{16 + 16 \log \frac{\mu(B(x,\Delta))}{\mu(B(x,\Delta/8))}}$$

Corallary 1

Corallary (Padded: Counting Measure)

For any metric space X, the Partition(Δ, X) algorithm gives an ε-padded decomposition bundle with

$$\varepsilon(x,\Delta) = \frac{1}{16 + 16\log\frac{|B(x,\Delta)|}{|B(x,\Delta/8)|}}$$

Corallary 2

Lemma (Doubling Constant)

For any metric space X, there exists measure μ^* such that for all $x \in X$ and all $\Delta > 0$:

$$\frac{\mu^*(B(x,\Delta))}{\mu^*(B(x,\Delta/2))} \leq \lambda_X$$

Corallary (Padded: Doubling Constant)

Using the μ^* from above we get an $\varepsilon(x,\Delta)$-padded decomposition bundle with:

$$\varepsilon(x,\Delta) = \mathcal{O}\left(\frac{1}{\log \lambda_X}\right)$$

Main Embedding Lemma

Lemma (Main Embedding Lemma)

Let μ be a non-degenerate measure on X, and let $\{P_u\}$ be an $\varepsilon(x,\Delta)$-padded decomposition bundle on X. Then there exists map $\phi : X \to L_2$ such that $\forall x, y \in X$:

$$\sqrt{V_\mu(x,y)} \cdot \min\{\delta_\varepsilon(x,y), \delta_\varepsilon(y,x)\} \le \frac{\|\phi(x) - \phi(y)\|_2}{d(x,y)} \le C\sqrt{\log \Phi(\mu)}$$

$$V_\mu(x,y) = \max\left[\log \frac{\mu(B(x, 2d(x,y)))}{\mu(B(x, d(x,y)/512))}, \log \frac{\mu(B(y, 2d(x,y)))}{\mu(B(y, d(x,y)/512))}\right]$$

$$\delta_\varepsilon(x,y) = \min\left[\varepsilon(x, 2^u) : u \in \mathbb{Z} \text{ and } \frac{d(x,y)}{32} \le 2^u \le \frac{d(x,y)}{2}\right]$$

$$\Phi(\mu) = \max_{x \in X} \frac{\mu(X)}{\mu(x)}$$

Proof of Main Thm

Theorem (Main Thm L_2)

Every n-point metric space (X, d) with doubling constant λ_x can be embedded in L_2 with distortion:

$$\mathcal{O}(\sqrt{\log \lambda_X \log n})$$

- Let μ be the counting measure.
- Concatenate two embeddings (one from each padded corallary)

Proof of Main Thm

Theorem (Main Thm L_2)

Every n-point metric space (X, d) with doubling constant λ_x can be embedded in L_2 with distortion:

$$\mathcal{O}(\sqrt{\log \lambda_X \log n})$$

- Let μ be the counting measure.
- Concatenate two embeddings (one from each padded corallary)

Proof (cont.)

$$\frac{\sqrt{V_\mu(x,y)}}{16+16V_\mu(x,y)} \leq \frac{\|\phi_1(x)-\phi_1(y)\|_2}{d(x,y)} \leq C\sqrt{\log n}$$

$$\frac{\sqrt{V_\mu(x,y)}}{\log \lambda_X} \leq \frac{\|\phi_2(x)-\phi_2(y)\|_2}{d(x,y)} \leq C\sqrt{\log n}$$

$$\frac{\|\phi(x)-\phi(y)\|_2^2}{d(x,y)^2} \geq \frac{V_\mu(x,y)}{[16+16V_\mu(x,y)]^2} + \frac{V_\mu(x,y)}{[\log \lambda_X]^2} \geq \Omega\left(\frac{1}{\log \lambda_X}\right)$$

Proof (cont.)

$$\frac{\sqrt{V_\mu(x,y)}}{16+16V_\mu(x,y)} \le \frac{\|\phi_1(x)-\phi_1(y)\|_2}{d(x,y)} \le C\sqrt{\log n}$$

$$\frac{\sqrt{V_\mu(x,y)}}{\log \lambda_X} \le \frac{\|\phi_2(x)-\phi_2(y)\|_2}{d(x,y)} \le C\sqrt{\log n}$$

$$\frac{\|\phi(x)-\phi(y)\|_2^2}{d(x,y)^2} \ge \frac{V_\mu(x,y)}{[16+16V_\mu(x,y)]^2} + \frac{V_\mu(x,y)}{[\log \lambda_X]^2} \ge \Omega\left(\frac{1}{\log \lambda_X}\right)$$

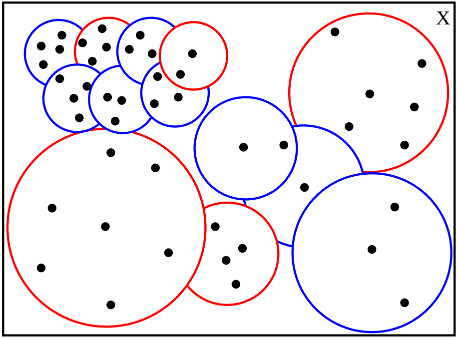
Proof (cont.)

$$\frac{\sqrt{V_\mu(x,y)}}{16+16V_\mu(x,y)} \leq \frac{\|\phi_1(x)-\phi_1(y)\|_2}{d(x,y)} \leq C\sqrt{\log n}$$

$$\frac{\sqrt{V_\mu(x,y)}}{\log \lambda_X} \leq \frac{\|\phi_2(x)-\phi_2(y)\|_2}{d(x,y)} \leq C\sqrt{\log n}$$

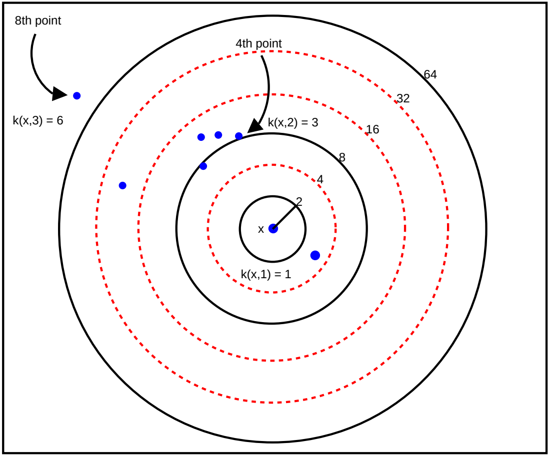
$$\frac{\|\phi(x)-\phi(y)\|_2^2}{d(x,y)^2} \geq \frac{V_\mu(x,y)}{[16+16V_\mu(x,y)]^2} + \frac{V_\mu(x,y)}{[\log \lambda_X]^2} \geq \Omega\left(\frac{1}{\log \lambda_X}\right)$$

Main Idea of Embedding



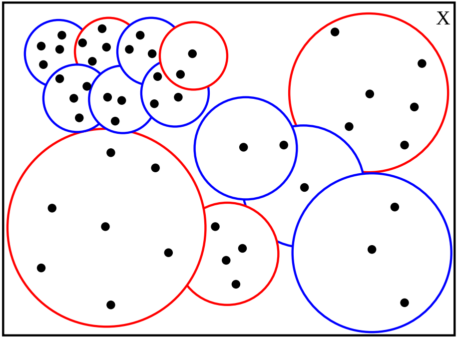
- Each coordinate will be distance to a set W_t that is composed of many different scales "glued" together.

The Embedding: Kappa



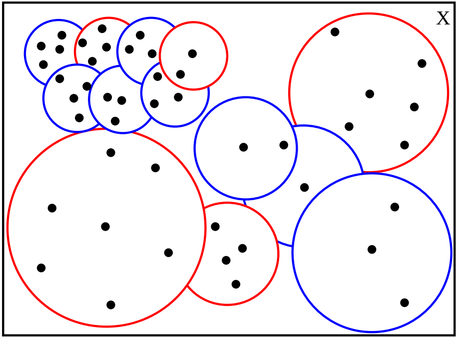
- $\kappa(x,t) = \max\{\kappa \in \mathbb{Z} : |B(x, 2^\kappa)| < 2^t\}$

The Embedding



- $T = \{1, \ldots, \log n\}$
- For every $C \subseteq X$ let $\sigma_u(C) \in \{0, 1\}$ u.a.r.
- Each x now "sniffs" around it's neighborhood to determine membership in W_t.
- Let $W_t = \{x \in X : \sigma_{\kappa(x,t)}(P_{\kappa(x,t)}(x)) = 0\}$
- $f(x) = (d(x, W_t) : t \in T)$.

The Embedding



- $T = \{1, \ldots, \log n\}$
- For every $C \subseteq X$ let $\sigma_u(C) \in \{0,1\}$ u.a.r.
- Each x now "sniffs" around it's neighborhood to determine membership in W_t.
- Let $W_t = \{x \in X : \sigma_{\kappa(x,t)}(P_{\kappa(x,t)}(x)) = 0\}$
- $f(x) = (d(x, W_t) : t \in T)$.

Proof Idea for Main Embedding Lemma

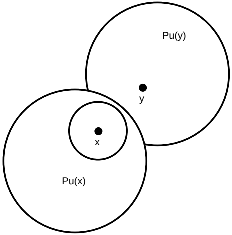
- Fix $x, y \in X$. Upper bound is easy since ϕ is coordinate-wise non-expanding.
- Lower bound we only show in expectation.
- We could apply a Chernoff style argument to say this bound holds with high probabilty by repeating the embedding $\mathcal{O}(\log n)$ times.
    - Giving embedding dimension $\mathcal{O}(\log^2 n)$.

Proof Idea for Main Embedding Lemma

- Fix $x, y \in X$. Upper bound is easy since ϕ is coordinate-wise non-expanding.
- Lower bound we only show in expectation.
- We could apply a Chernoff style argument to say this bound holds with high probabilty by repeating the embedding $\mathcal{O}(\log n)$ times.
  - Giving embedding dimension $\mathcal{O}(\log^2 n)$.

Lower Bound

- Consider $x, y \in X$ where $d(x,y) \approx 2^m$.



1. Padded Decomposition Property
   $\rightarrow Pr[B(x, \varepsilon(x,\Delta)\Delta) \subseteq P_u(x)] \geq 1/2$
2. $Pr(x \notin W_t \ \wedge \ y \in W_t) = (\frac{1}{2})(\frac{1}{2}) = 1/4$
3. Conclusion: $d(x, W_t) = \Omega(2^m) \approx \Omega(d(x,y))$ (with prob. $\geq \alpha$).
4. Above holds for each t where $\kappa(x,t) \approx m$.

Some subtleties...

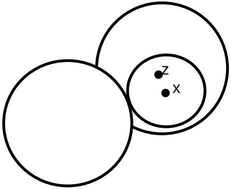
- There is one subtlety that we glossed over in the discussion above...

Subtlety:

- Padded Decomposition Property doesn't strictly apply to W_t. Why?

$$Pr\left[B\left(x, \varepsilon 2^m\right) \subseteq P_m(x)\right] \geq 1/2$$

- Consider $z \in B(x, 2^m)$ – Whose to say that both z and x "sniffed" at the same scale?
- We need a smoothness property of $\kappa(x, t)$.



Some subtleties...

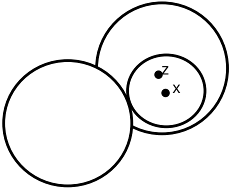
- There is one subtlety that we glossed over in the discussion above...

Subtlety:

- Padded Decomposition Property doesn't strictly apply to W_t. Why?

$$Pr\left[B\left(x, \varepsilon 2^m\right) \subseteq P_m(x)\right] \geq 1/2$$

- Consider $z \in B(x, 2^m)$ – Whose to say that both z and x "sniffed" at the same scale?
- We need a smoothness property of $\kappa(x, t)$.



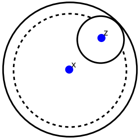
Smoothness of $\kappa(x,t)$

Lemma

Let $x \in X$ and $m = \kappa(x,t)$. Let $z \in B(x,\tau)$, where $\tau \le \frac{2^m}{10}$. Let $m' = \kappa(z,t)$, then:

$$m' \in \{m-4, m-3, m-2, m-1, m, m+1\}$$

1. $|B(x,2^m)| \approx 2^t$ and $d(x,z) \le 2^m$.
2. So, $B(x,2^m) \subset B(z,2^m+2^m) = B(z,2^{m+1})$. There must be at least 2^t points in $B(z,2^{m+1})$ so $m' \le m+1$.
3. Similarly $|B(z,2^{m'})| \approx 2^t$ and $d(x,z) \le \frac{2^m}{10}$. Therefore, $B(z,2^{m'}) \subset B(x,2^{m'}+\frac{2^m}{10})$ and $B(x,2^m) \subset B(x,2^{m'}+\frac{2^m}{10})$. So, $2^m \le 2^{m'}+\frac{2^m}{10} \Rightarrow m' \ge m-4$.



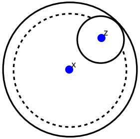
Smoothness of $\kappa(x,t)$

Lemma

Let $x \in X$ and $m = \kappa(x,t)$. Let $z \in B(x,\tau)$, where $\tau \leq \frac{2^m}{10}$. Let $m' = \kappa(z,t)$, then:

$$m' \in \{m-4, m-3, m-2, m-1, m, m+1\}$$

1. $|B(x,2^m)| \approx 2^t$ and $d(x,z) \leq 2^m$.
2. So, $B(x,2^m) \subset B(z,2^m+2^m) = B(z,2^{m+1})$. There must be at least 2^t points in $B(z,2^{m+1})$ so $m' \leq m+1$.
3. Similarly $|B(z,2^{m'})| \approx 2^t$ and $d(x,z) \leq \frac{2^m}{10}$. Therefore, $B(z,2^{m'}) \subset B(x,2^{m'}+\frac{2^m}{10})$ and $B(x,2^m) \subset B(x,2^{m'}+\frac{2^m}{10})$. So, $2^m \leq 2^{m'}+\frac{2^m}{10} \Rightarrow m' \geq m-4$.



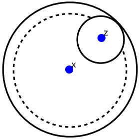
Smoothness of $\kappa(x, t)$

Lemma

Let $x \in X$ and $m = \kappa(x, t)$. Let $z \in B(x, \tau)$, where $\tau \leq \frac{2^m}{10}$. Let $m' = \kappa(z, t)$, then:

$$m' \in \{m-4, m-3, m-2, m-1, m, m+1\}$$

1. $|B(x, 2^m)| \approx 2^t$ and $d(x, z) \leq 2^m$.
2. So, $B(x, 2^m) \subset B(z, 2^m + 2^m) = B(z, 2^{m+1})$. There must be at least 2^t points in $B(z, 2^{m+1})$ so $m' \leq m + 1$.
3. Similarly $|B(z, 2^{m'})| \approx 2^t$ and $d(x, z) \leq \frac{2^m}{10}$. Therefore, $B(z, 2^{m'}) \subset B(x, 2^{m'} + \frac{2^m}{10})$ and $B(x, 2^m) \subset B(x, 2^{m'} + \frac{2^m}{10})$. So, $2^m \leq 2^{m'} + \frac{2^m}{10} \Rightarrow m' \geq m - 4$.



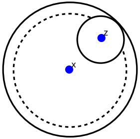
Smoothness of $\kappa(x,t)$

Lemma

Let $x \in X$ and $m = \kappa(x,t)$. Let $z \in B(x,\tau)$, where $\tau \leq \frac{2^m}{10}$. Let $m' = \kappa(z,t)$, then:

$$m' \in \{m-4, m-3, m-2, m-1, m, m+1\}$$

1. $|B(x,2^m)| \approx 2^t$ and $d(x,z) \leq 2^m$.
2. So, $B(x,2^m) \subset B(z,2^m+2^m) = B(z,2^{m+1})$. There must be at least 2^t points in $B(z,2^{m+1})$ so $m' \leq m+1$.
3. Similarly $|B(z,2^{m'})| \approx 2^t$ and $d(x,z) \leq \frac{2^m}{10}$. Therefore, $B(z,2^{m'}) \subset B(x,2^{m'}+\frac{2^m}{10})$ and $B(x,2^m) \subset B(x,2^{m'}+\frac{2^m}{10})$. So, $2^m \leq 2^{m'}+\frac{2^m}{10} \Rightarrow m' \geq m-4$.



How to apply Padded Decomposition

- So, there are only a finite number of different scales to worry about.
- $\Pr(x \text{ is padded at all 6 scales}) \geq 2^{-6}$.
- So the prob. of our "good" event is still bounded away from zero.

Other contributions of KMLN04

- A stronger notion of distortion would consider k-tuples of points instead of pairwise distances.
- This same embedding is (k, $\mathcal{O}(\sqrt{\log \lambda_X \log n})$) - volume respecting in ℓ_2.
  - Immediately gives $\mathcal{O}(r\sqrt{\log n})$ approx min-cut alg for graphs excluding $K_{r,r}$.
- A slight variation gives $\mathcal{O}(1)$ distortion of planar metrics into $\ell_\infty^{\mathcal{O}(\log n)}$.

Other contributions of KMLN04

- A stronger notion of distortion would consider k-tuples of points instead of pairwise distances.
- This same embedding is (k, $\mathcal{O}(\sqrt{\log \lambda_X \log n})$) - volume respecting in ℓ_2.
  - Immediately gives $\mathcal{O}(r\sqrt{\log n})$ approx min-cut alg for graphs excluding $K_{r,r}$.
- A slight variation gives $\mathcal{O}(1)$ distortion of planar metrics into $\ell_\infty^{\mathcal{O}(\log n)}$.

Other contributions of KMLN04

- A stronger notion of distortion would consider k-tuples of points instead of pairwise distances.
- This same embedding is (k, $\mathcal{O}(\sqrt{\log \lambda_X \log n})$) - volume respecting in ℓ_2.
  - Immediately gives $\mathcal{O}(r\sqrt{\log n})$ approx min-cut alg for graphs excluding $K_{r,r}$.
- A slight variation gives $\mathcal{O}(1)$ distortion of planar metrics into $\ell_\infty^{\mathcal{O}(\log n)}$.