Dimensionality Reduction:
Theoretical Analysis of Practical Measures

Nova Fandina
Hebrew University

Joint work with Yair Bartal, Hebrew University
Ofer Neiman, Ben Gurion University

Outline

- Measuring the Quality of Embedding

  - in theory : worst case distortion analysis
  - in practice: average case distortion measures
  - in between: theoretical analysis of practical measures (for dimensionality reduction methods)

- Our Results

  - upper bounds
  - lower bounds
  - approximating optimal embedding

Measuring the Quality of Embedding: in theory

Basic question in metric embedding theory (informally)

Given metric spaces X and Y, embed X into Y,
with small error on the distances

How *well* it can be done?

In theory: “well” traditionally means to minimize **distortion** of the **worst** pair

Definition

For an embedding $f: X \to Y$ *, for a pair of points* $u \neq v \in X$

- $expans_f(u,v) = \frac{d_Y(f(u),f(v))}{d_X(u,v)}$, $contr_f(u,v) = \frac{d_X(u,v)}{d_Y(f(u),f(v))}$
- $distortion(f) = max_{u \neq v \in X}\{expans_f(u,v)\} \cdot max_{u \neq v \in X}\{contr_f(u,v)\}$

Measuring the Quality of Embedding: in practice

Demand for the worst case guarantee is too strong:
The quality of a method in practical applications is its average performance over all pairs

- Yuval Shavitt and Tomer Tankel. **Big-bang simulation for embedding network distances in Euclidean space**. IEEE/ACM Trans. Netw., 12(6), 2004.
- P. Sharma, Z. Xu, S. Banerjee, and S. Lee**. Estimating network proximity and latency.** Computer Communication Review, 36(3), 2006.
- P. J. F. Groenen, R. Mathar, and W. J. Heiser. **The majorization approach to multidimensional scaling for minkowski distances**. Journal of Classification, 12(1), 1995.
- J. F. Vera, W. J. Heiser, and A. Murillo. **Global optimization in any minkowski metric: A permutation-translation simulated annealing algorithm for multidimensional scaling**. Journal of Classification, 24(2), 2007.
- A. Censi and D. Scaramuzza. **Calibration by correlation using metric embedding from nonmetric similarities.** IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(10), 2013.
- C. Lumezanu and N. Spring**. Measurement manipulation and space selection in network coordinates**. The 28th International Conference on Distributed Computing Systems, 2008.
- S. Chatterjee, B. Neff, and P. Kumar. **Instant approximate 1-center on road networks via embeddings.** In Proceedings of the 19th International Conference on Advances in Geographic Information Systems, GIS '11, 2011.
- S. Lee, Z. Zhang, S. Sahu, and D. Saha. **On suitability of Euclidean embedding for host-based network coordinate systems**. IEEE/ACM Trans. Netw., 18(1), 2010.
- L. Chennuru Vankadara and U. von Luxburg. **Measures of distortion for machine learning**. Advances in Neural Information Processing Systems, Curran Associates, Inc., 2018.

Just a small sample from googolplex number of such studies

Measuring the Quality of Embedding: in practice

Moments of Distortion and Relative Error

For $f: X \to Y$, for a pair $u \neq v \in X$, $dist_f(u,v) \coloneqq \max\{expans_f(u,v), contract_f(u,v)\}$

ℓ_q-distortion ABN[06]

For $f: X \to Y$, for a distribution Π over pairs of X, $q \geq 1$

$$\ell_q^{(\Pi)}\text{-}dist(f) = \left(E_\Pi\left[\left(dist_f(u,v)\right)^q\right]\right)^{1/q}$$

Relative Error Measure [In many papers]

$$REM_q^{(\Pi)} = \left(E_\Pi\left[\left(|dist_f(u,v) - 1|\right)^q\right]\right)^{1/q}$$

Measuring the Quality of Embedding: in practice

Additive Distortion Measures

[MDS: optimally embed a given finite X into a k-dim Euclidean space, for a given k]

For a pair $u \neq v \in X$, $d_{uv} = d_X(u,v)$, $\hat{d}_{uv} = d_Y\left(f(u), f(v)\right)$

$$Stress_q(f) = \left(\frac{E_\Pi[|d_{uv} - \hat{d}_{uv}|^q]}{E_\Pi[(d_{uv})^q]}\right)^{1/q}$$

$$Stress^*_q(f) = \left(\frac{E_\Pi[|d_{uv} - \hat{d}_{uv}|^q]}{E_\Pi[(\hat{d}_{uv})^q]}\right)^{1/q}$$

$$Energy_q(f) = \left(\mathrm{E}_\Pi\left[\left(\frac{|\hat{d}_{uv} - d_{uv}|}{d_{uv}}\right)^q\right]\right)^{1/q}$$

$$REM_q(f) = \left(\mathrm{E}_\Pi\left[\left(\frac{|\hat{d}_{uv} - d_{uv}|}{\min\{d_{uv}, \hat{d}_{uv}\}}\right)^q\right]\right)^{1/q}$$

Measuring the Quality of Embedding: in practice

σ-distortion ML motivated, in [VvL18]

$$\sigma - dist^{(\Pi)}{}_{q,r}(f) = \left(\mathrm{E}_{\Pi} \left[\left(\left| \frac{exapns_f(u,v)}{\ell_r^U - expans(f)} - 1 \right| \right)^q \right] \right)^{1/q}$$

- $\ell_r^{(\mathrm{U})} - expans(f) = E_{\mathrm{U}}[(expans_f(u,v)^r)]$
- $\ell_r^{(\mathrm{U})} - contr(f) = E_{\mathrm{U}}[(contr_f(u,v)^r)]$

"Necessary properties for ML applications"

- translation invariance
- scale invariance
- monotonicity
- robustness (outliers, noise)
- incorporation of probability

- Many heuristics for optimizing these measures
- Almost nothing is known in terms of rigorous analysis

Measuring the Quality of Embedding: in between

Bridging the gap between theory and practice outlook

$\alpha(k, q)$-Dimension Reduction

Given a dimension bound $\mathbf{k} \geq \mathbf{1}$ and $\mathbf{q} \geq \mathbf{1}$, what is the least $\boldsymbol{\alpha}(\mathbf{k}, \mathbf{q})$ such that every finite subset of Euclidean space embeds into $\mathbf{k}$ dim. with $\mathbf{Measure_q} \leq \boldsymbol{\alpha}(\mathbf{k}, \mathbf{q})$**?**

General Metrics: Approximating the Optimal Embedding

For a given finite X and for $k \geq 1$, compute an embedding of X into k-dim Euclidean space that *approximates* the best possible embedding, for a given $Measure_q$.

[CD06] Optimizing is NP-hard for $Stress_q$ and $k = 1$

Our Results: upper bounds

previous results

$\alpha(k,q)$-Dimension Reduction

Given a dimension bound $\mathbf{k} \geq \mathbf{1}$ and $\mathbf{q} \geq \mathbf{1}$, what is the least $\boldsymbol{\alpha}(\mathbf{k},\mathbf{q})$ such that every finite subset of Euclidean space embeds into $\mathbf{k}$ dim. with $\mathbf{Measure_q} \leq \boldsymbol{\alpha}(\mathbf{k},\mathbf{q})$?

Previous results: *worst case distortion*

JL[84] Every n-point $X \in \ell_2^d$ embeds into ℓ_2^k with **distortion** $O\left(n^{\frac{2}{k}}\sqrt{(\log n)/k}\right)$

Mat[90] There is $V \in \ell_2^{k+1}$ such that any $f: V \to \ell_2^k$ must have **distortion** $n^{\Omega(1/k)}$

- distortion(f) $\leq$ $(\ell_\infty\text{-dist})^2$
- For every $f: X \to Y$ (scalable) there is g: $X \to Y$ with ℓ_∞**-dist(g)** $= \sqrt{\mathbf{dist(f)}}$

What about the $Measure_q$ guarantees for $q < \infty$?

Our Results: upper bounds

JL transform: IM implementation

The answer to the $\alpha(k,q)$-Dim. Reduction question is, essentially, the JL transform

[JL84] Projection onto a random subspace of dim. $\boldsymbol{k = O(\log n / \epsilon^2)}$,
with const. prob. $\boldsymbol{dist(f) = 1 + \epsilon}$ [tight, LN16]

[IM 98] T is a matrix of size $k \times d$ with indep. entries sampled from $N(0,1)$.
The embedding $f: X \to \ell_2^k$ is $f(x) = 1/\sqrt{k} \cdot T(x)$

- The JL transform of IM98 provides constant upper bounds for all $Measure_q$
  The bounds are almost optimal
- Other popular implementations of JL do not work for ℓ_q-dist and for REM_q
- PCA may produce an embedding of extremely poor quality for all the measures (this does not happen to the JL)

Our Results: upper bounds other implementations of JL

[Achl03] The entries of T are uniform indep. from $\{\pm 1\}$

[DKS10,KN10, AL10] Sparse/Fast: particular distr. from $\{\pm 1,0\}$

Constant bounds cannot be achieved using the above implementations

Observation

If a linear transformation $T: R^d \to R^k$ samples its entries form a discrete set of values of size $s \le d^{1/k}$, then applying it on a standard basis of R^d results in ℓ_q-dist, $REM_q = \infty$.

- $\ell_q^{(\Pi)}\text{-}dist(f) = \left(E_\Pi\left[\left(dist_f(u,v)\right)^q\right]\right)^{1/q}$, $REM_q^{(\Pi)} = \left(E_\Pi\left[\left(|dist_f(u,v) - 1|\right)^q\right]\right)^{1/q}$
- $dist_f(u,v) = \max(expans_f(u,v), contract_f(u,v))$

➢ $T(e_1, \dots, e_d) = \{columns\ of\ T\}$. The number of different columns is $s^k < d$

Our Results: upper bounds

limitation of PCA

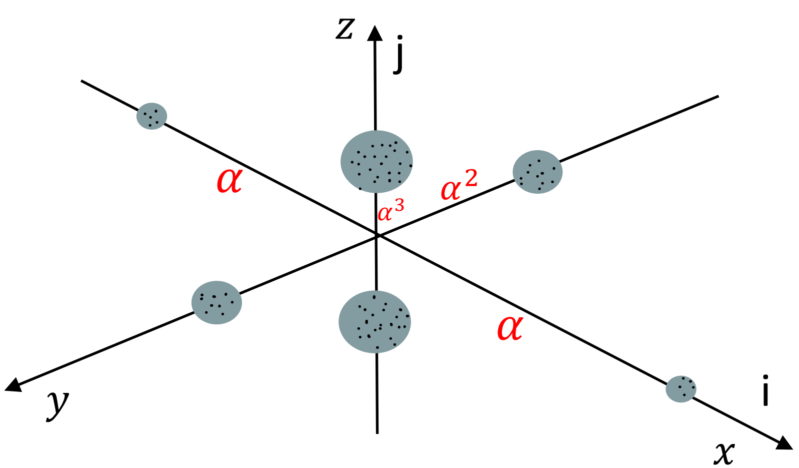
PCA/c-MDS For a given finite $X \in \ell_2^d$ and a given integer $k \geq 1$, computes the best rank k- approx. to X:

A projection P onto the k- dim subspace spanned by largest eigenvectors of the covariance matrix, with the smallest $\sum_{u \in X} \|u - P(u)\|^2$

- $f: X \to \ell_2^k$ with *optimal* $\sum_{u \neq v \in X} (d_{uv}^2 - \hat{d}_{uv}^2)$ over all *projections*
- Often misused: "minimizing $Stress_2$ over all embeddings into k-dim"
- Actually, PCA does not minimize any of the mentioned measures

Our Results: upper bounds

- The metric is in d dimensional Euclidean space, for any d large enough
- Fix some $\alpha < 1$, and $q \geq 1$
- Consider the standard basis vectors $e_1, \dots, e_d$
- For each vector e_i, let X_i be the set of $\left(\frac{1}{\alpha^i}\right)^q$ copies of vector $\alpha^i \cdot e_i$, and let Y_i be the set of the same size of the antipodal vector $-\alpha^i \cdot e_i$



- $Stress_2^2$ measure: $\frac{\sum_{u \neq v \in X} \left(d_{uv} - \hat{d}_{uv}\right)^2}{\sum_{u \neq v \in X} d_{uv}^2}$
- $\sum_{u \neq v} d_{uv}^2 \approx d \cdot \left(\frac{1}{\alpha^d}\right)^2$
- pairs between X_i and X_j, $i < j$ contribute:

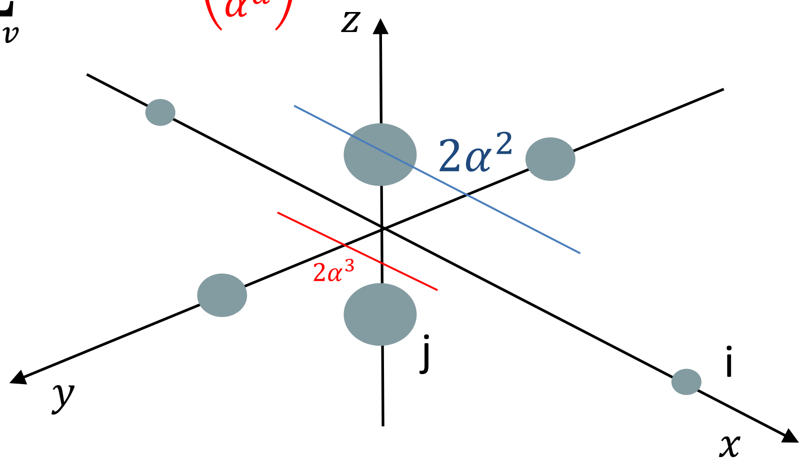
$$\frac{1}{\alpha^{2i}} \cdot \frac{1}{\alpha^{2j}} \cdot (\alpha^{2i} + \alpha^{2j}) \leq \frac{1}{\alpha^{2i}} \cdot \frac{1}{\alpha^{2j}} \cdot 2\alpha^{2i} \leq 2\frac{1}{\alpha^{2j}}$$

$$\frac{1}{\alpha^{2i}} \cdot \frac{1}{\alpha^{2j}} \cdot (\alpha^{2i} + \alpha^{2j}) \geq \frac{1}{\alpha^{2j}}$$

Our Results: upper bounds

limitation of PCA

$$\sum_{u \neq v} d_{uv}^2 \approx d \cdot \left(\frac{1}{\alpha^d}\right)^2$$



- Error contribution: $i < j$
- $\approx \frac{1}{\alpha^{2i}} \cdot \frac{1}{\alpha^{2j}} \left(\sqrt{(\alpha^{2i} + \alpha^{2j})} - (\alpha^i - \alpha^j) \right)^2$

$\approx 1/\alpha^{2i}$

for $i < j$, in total $\approx \left(\frac{1}{\alpha^{d-1}}\right)^2$

- $\boldsymbol{Stress_2 \leq \alpha/d^{1/2}}$

➢ PCA projects onto $span\{e_1, e_2, \dots, e_k\}$ taking $k < 0.99d$

- Error contribution: $\geq (d-k)\left(1/\alpha^d\right)^2$
- $Stress_2 \geq \Omega(1)$
- Is not better than a naïve algo: any non-expansive embedding

- ℓ_q-dist/$REM_q = \infty$

➢ The JL embedding has bounded measures: $\alpha(k) \to 0$, as k increases

Our Results: upper bounds — moment analysis of JL transform

Theorem (Moment analysis of JL transform)

There is a map (JL or normalized JL) $f: X \to \ell_2^k$ s.t. for a given $q \geq 1$ with const. prob.

	$1 \leq q < \sqrt{k}$	$\sqrt{k} \leq q \leq k/4$	$k/4 \leq q \leq k$	$q = k$	$k \leq q \leq \infty$
ℓ_q-dist(f) $=$	$1 + O\left(\frac{1}{\sqrt{k}}\right)$	$1 + O\left(\frac{q}{k-q}\right)$	$\left(\frac{k}{k-q}\right)^{O(1/q)}$	$O\left(\sqrt{\log n}\right)^{1/k}$	$n^{O\left(\frac{1}{k}-\frac{1}{q}\right)}$

The bounds are almost tight in most of the ranges of values of q

Proof (for q < k) For a given dist. Π over pairs of X, for a given $q \geq 1$

$$E_f\left[\left(\ell_q^{(\Pi)} - dist(f)\right)^q\right] = E_f\left[E_{(u,v)\sim\Pi}\left(dist_f(u,v)\right)^q\right] = E_{(u,v)\sim\Pi}\left[E_f\left[\left(dist_f(u,v)\right)^q\right]\right]$$

For every $u \neq v \in X$, estimate $E_f\left[\left(dist_f(u,v)\right)^q\right] =$

$$E_f\left[\left(\max\left(\frac{\|f(u)-f(v)\|}{\|u-v\|}, \frac{\|u-v\|}{\|f(u)-f(v)\|}\right)\right)^q\right]$$

Since f is a linear map, for any $z \in \mathbb{R}^d$, with $\|z\| = 1$ estimate

$$E_f\left[\max\left(\|f(z)\|^q, \frac{1}{\|f(z)\|^q}\right)\right]$$

- $f(z) = 1/\sqrt{k} \cdot T(z) = 1/\sqrt{k} \cdot (< z, T_1 >, \dots, < z, T_k >) = 1/\sqrt{k}\,(Y_1, \dots, Y_k)$, for $Y_i \sim N(0,1)$
- $\|f(z)\|^q = (\|f(z)\|^2)^{q/2} = \left(\frac{X}{k}\right)^{q/2}$, where $X \sim \chi_k^2$

- $E_f\left[\max\left(\|f(z)\|^q, \frac{1}{\|f(z)\|^q}\right)\right] \leq E_{X\sim\chi_k^2}[(X/k)^{q/2}] + E_{X\sim\chi_k^2}[(k/X)^{q/2}]$

goes to ∞ , as $q \to k$

- $E_{X\sim\chi_k^2}[(k/X)^{q/2}] = \int_0^\infty \left(\frac{k}{x}\right)^{\frac{q}{2}} \frac{x^{\frac{k}{2}-1}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)e^{\frac{x}{2}}}\,dx = \frac{\left(\frac{k}{2}\right)^{\frac{q}{2}}\Gamma\left(\frac{k}{2}-\frac{q}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$

Gamma function: $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx,\ \Gamma(t+1) = t\,\Gamma(t)$

- $E_f\left[\left(\ell_q - contr(f)\right)^q\right] = \left(1 + O\left(\frac{q}{k-q}\right)\right)^q$, for all $q < k$
- $E_f\left[\left(\ell_q - expans(f)\right)^q\right] = \left(1 + O\left(\frac{q}{k}\right)\right)^q$, for all $q \geq 1$

- $E_f[\ell_q - dist(f)] \leq 2^{\frac{1}{q}}\left(1 + O\left(\frac{q}{k-q}\right)\right) \leq \left(1 + \frac{1}{q}\right)\left(1 + O\left(\frac{q}{k-q}\right)\right)$

$= 1 + O(1/\sqrt{k})$, taking $q = \sqrt{k}$ ■

Our Results: upper bounds

estimates for large values of q

Theorem (Moment analysis of JL transform)

There is a map (normalized JL) $f: X \to \ell_2^k$ s.t. for a given $q \geq 1$ with const. prob.

	$1 \leq q < \sqrt{k}$	$\sqrt{k} \leq q \leq k/4$	$k/4 \leq q \leq k$	$q = k$	$k \leq q \leq \infty$
ℓ_q-dist(f) $=$	$1 + O\left(\frac{1}{\sqrt{k}}\right)$	$1 + O\left(\frac{q}{k-q}\right)$	$\left(\frac{k}{k-q}\right)^{O(1/q)}$	$O\left(\sqrt{\log n}\right)^{1/k}$	$n^{O\left(\frac{1}{k}-\frac{1}{q}\right)}$

Proof The same as before, estimation of expectation conditioning on the event

$\forall\, u \neq v \in X,\ contr_f(u,v) \leq n^{\frac{2}{k}}$ (holds with const prob).

Normalize by an appropriate factor (depends on q and k)

■

Our Results: upper bounds

Theorem (REM and additive measures analysis of JL)

There is a map (JL) $f: X \to R^k$, for $k \geq 2$, s.t. with const. prob.
for all $1 \leq q \leq k-1$ (simultaneously)

$$\sigma\text{-dist}, Stress_q, Stress^*, Eneregy_q(f) \leq REM_q(f) = O(\sqrt{q/k})$$

- All the additive measures $\leq\ REM$

$$REM_q = E_\Pi[|dist_f(u,v) - 1|^q]$$

- As before, estimate the appropriate integral to get the bound

- Tight for $q \geq 2$ (equilateral space on n points, E_n)
  Based on [Alon90]: lower bound for embedding with $1+\epsilon$ w.c. dist
  For $1 \leq q < 2$ the lower bound is $\Omega(1/k^{1/q})$

Our Results: optimal embedding for E_n small q

- There is a gap with respect to the upper bounds provided by the JL embedding

Theorem

➢ Any embedding $f: E_n \to \ell_2^k$ must have:

$$\ell_q\text{- dist}(f) = 1 + \Omega(q/k) \qquad \text{for } 1 \le q \le \sqrt{k}$$
$$Energy_q(f) = \Omega\left((1/k)^{1/q}\right) \qquad \text{for } 1 \le q < 2$$

- This is the best we can do for E_n space:

For every $k \ge 3$, for all $1 \le q \le \sqrt{k}$, for every distr. Π over pairs of E_n,

there is a random map $f: E_n \to \ell_2^k$ s.t. with const. prob.

$$\ell_q\text{ - dist}(f) = 1 + O(q/k)$$

Our Results: optimal embedding for E_n

small $q < \sqrt{k}$

Theorem

For every $k \geq 3$, for all $1 \leq q \leq \sqrt{k}$, for every distr. Π over pairs of E_n, there is a random

$f: E_n \to \ell_2^k$ s.t. with const. prob. ℓ_q - dist$(f) = 1 + O(q/k)$

Intuition

uniform metric in $\boldsymbol{n}$ dim

$E_n = \{v_1, v_2, \dots, v_n\}$

v_i → uniformly chosen e_j

uniform metric in $\mathbf{k}$ dim

e_1 e_2 e_3 … e_k

v_3 v_1 v_2

v_4

$E(\ell_q\text{– dist}(f))^q = E\,[(dits_f(u,v))^q] = \Pr[f(u) = f(v)]$ ∞

$+ \Pr[f(u) \neq f(v)] \cdot 1$

Our Results: optimal embedding for E_n small $q < \sqrt{k}$

Algorithm: map $F: E_n \to \ell_2^k$

$\forall\, v \in E_n$

- ind. & uniformly choose a sphere $S[j]$
- place v ind. & uniformly on $S[j]$;
  $F_j(v) \to \ell_2^{q^2}$
- $F(v) \to$

00000	00000	$F_j(v)$	00000

- each $S[j]$ has radius $r_j = 1/\sqrt{2}$
- Each sphere is of dim. d, $S[j] \subset \ell_2^{q^2}$
- # spheres k/q^2

- If $v \in S[j]$ and $u \in S[i]$ $\implies$ $\|F(v) - F(u)\|^2 = \|F_j(v)\|^2 - \|F_j(u)\|^2 = 1/2 + 1/2 = 1$
- $E(\ell_q\text{– dist}(f))^q = E\,[(dits_f(u,v))^q] = \quad q^2/k \quad \cdot\ 2\,(d/(d-q))^{q/2}$
  $+ \Pr[f(u) \neq f(v)] \cdot 1 \ \ \leq^{[d=q^2]}\ 1 + (q^2/k)\cdot const$
- Taking $1/q$ on both sides, ℓ_q - dist $(f) = 1 + O(q/k)$.

Our Results: lower bounds

phase transition at $q = k$

Theorem

- For any $k \geq 1$, any embedding $f: E_n \to \ell_2^k$ has ℓ_k-dist$(f) = \frac{\Omega\left(\sqrt{\log n}\right)^{\frac{1}{k}}}{k^{1/4}}$.
- For any q $> k \geq 1$, any embedding $f: E_n \to \ell_2^k$ has ℓ_q-dist$(f) = \Omega(n^{\frac{1}{2k}-\frac{1}{2q}})$.

Proof

- It is enough to show that for any non-expansive $f: E_n \to \ell_2^k$, ℓ_k-dist$(f) \geq \frac{\Omega(\log n)^{\frac{1}{k}}}{\sqrt{k}}$
- **Claim:** if for any non-expansive $F: E_n \to Y, \ell_k$-dist$(F) \geq D(k,n)$, then
  for any $f: E_n \to Y, \ell_k$-dist$(f) \geq const \cdot \sqrt{D(k,n)}$.
- Since $\ell_2^k \sim \ell_\infty^k$ with distortion $\sqrt{k}$, it is enough to prove for any non-expansive $f: E_n \to \ell_\infty^k$ has ℓ_k-dist$(f) \geq \Omega(\log n)^{1/k}$

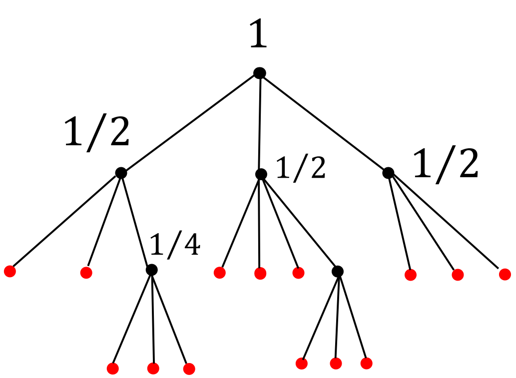
Our Results: (almost) optimal lower bound for $q > k$

Claim: For every non-expansive $f: E_n \to \ell_\infty^k$, ℓ_k-dist$(f) \geq \Omega(\log n)^{1/k}$.

Proof

- Basically, embedding (non-expansively) E_n into ℓ_∞^k is
  as embedding it (non-expansively) into a family of certain tree metrics

2-HST metrics of degree $\boldsymbol{k}$ − a family of all rooted trees on n leaves, with each node having at most 2^k children. The nodes have labels, decreasing by a factor of 2 along the paths from the root to each leaf. The root's label is 1.

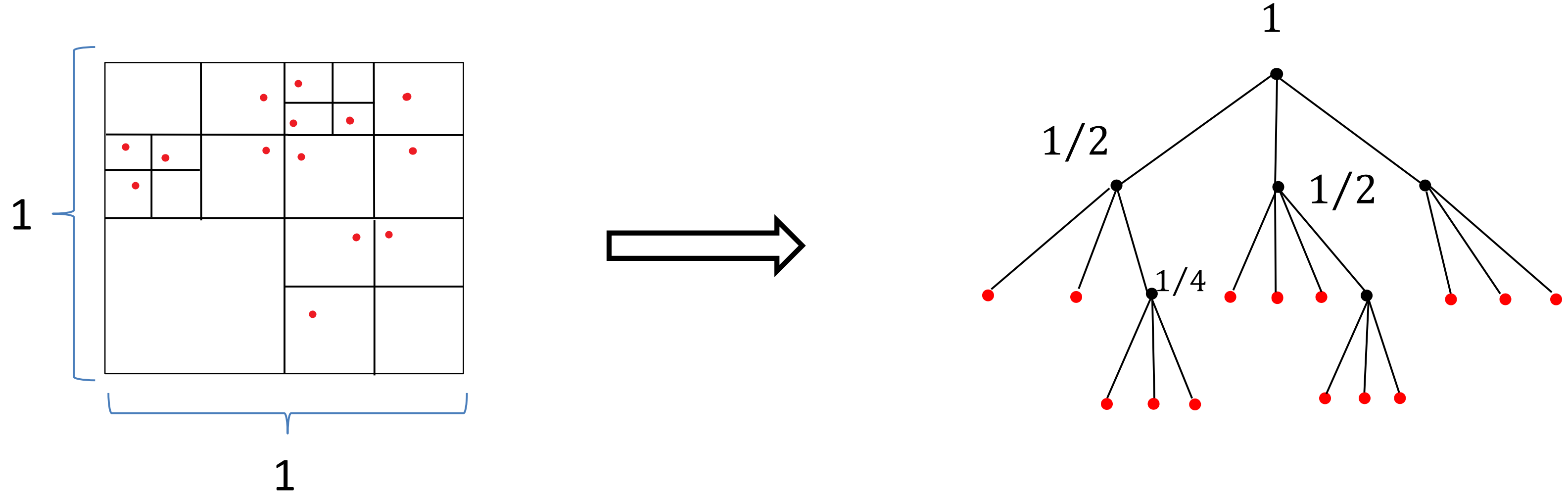


Each tree defines a metric over the set of its leaves: $dist(u, v) = label(lca(u, v))$

Our Results: (almost) optimal lower bound for $q > k$

- Every non-expansive embedding $f: E_n \to \ell_\infty^k$ can be modified to the one that embeds E_n into a 2-HST tree from the family of degree k, with better ℓ_k-distortion

For $k = 2$, r ecursively construct 2-HST tree, of degree 4



Our Results: (almost) optimal lower bound for $q = k$

So, it is enough to prove that:

Claim

Any non-expansive embedding f of E_n into a family of 2-HST's of degree k, has $(\ell_k\text{−}dist(f))^k \geq \Omega(\log n)$.

Proof By induction on n, showing that the best tree is the perfectly balanced (each node has exactly 2^k children). Computing its weight completes the proof. ■

Our Results: approximating optimal embedding

General Metrics: Approximating the Optimal Embedding

For a given finite X and for $k \geq 1$, compute an embedding of X into k-dim Euclidean space that *approximates* the best possible embedding, for a given $Measure_q$.

[ABN06] Every finite X embeds into $\ell_p^{O_p(\log n)}$, with ℓ_q-distortion $O(q/p)$.

Gives $O(q)$ approximation to the optimum under ℓ_q-distortion

[HIL03] 2-approx. to $Stress_\infty$, for embedding into $k = 1$ dim

[Bado03] $O(1)$-approx. to $Stress_\infty$, for embedding into $k = 2$ dim, under l_1 norm

[Dham04] $O(\log^{1/q} n)$-approx. to $Stress_q$, for embedding into $k = 1$ dim

Our Results: approximating optimal embedding

For a given X and q, $k \geq 1$, for an objective measure Obj_q :

$OPT: X \to \ell_2^k$ is an optimal embedding for the measure Obj_q

Theorem

For a given finite X, for a given $k \geq 3$ and $2 \leq q \leq k-1$, there is a randomized polytime algorithm that computes an embedding $\mathrm{F}: X \to \ell_2^k$ s.t. with const. prob.

- $l_q\text{-}dist(F) = \left(1 + O\left(\frac{\mathbf{1}}{\sqrt{\boldsymbol{k}}} + \frac{\boldsymbol{q}}{\boldsymbol{k-q}}\right)\right) \cdot \boldsymbol{OPT}$
- $Obj_q^{(\Pi)}(F) = O\left(Obj_q(\boldsymbol{OPT})\right) + O\left(\sqrt{q/k}\right)$, for all the rest objective measures

Proof Outline: (For Stress)

- Find an optimal embedding f without constraining the dimension
- Reduce dimension with JL (IM implementation) and show this works

Our Results: approximating optimal embedding   proof outline

Find an optimal embedding f without constraining the dimension

- [LLR95]: SDP computes a map with optimal worst case distortion

  For $X = \{x_0, \dots, x_{n-1}\}$, for each pair, variable $z_{ij} = \left\| f(x_i) - f(x_j) \right\|^2$,
  $g_{ij} = 1/2(z_{0i} + z_{0j} - z_{ij})$ represents $< f(x_i), f(x_j) >$, $f(x_0) = 0$
  $\sqrt{z_{ij}}$ are Euclidean distances iff matrix $G[i,j] \coloneqq g_{ij}$ is PSD

- Convex objective function for $Stress_q$: for $q \geq 2$

$$\min \quad \sum_{0<i<n} \left(\left(\sqrt{\frac{z_{ij}}{d_{ij}}} - 1 \right)^2 \right)^{q/2} \quad , \text{ s.t. } \quad z_{ij} \geq 0, G \text{ is PSD}$$

Our Results: approximating optimal embedding   proof outline

- Apply JL $g: f(X) \to \ell_2^k$
  - $F \coloneqq g \circ f$

Claim

If $f: X \to Y$ is some embedding and $g: Y \to Z$ is a random map that has $E\left[expans_g(u,v)\right] = A$, and $E\left[contr_g(u,v)\right] = B$, for all pairs in X, then

$$E\left[Energy_q(g \circ f)\right] \leq 4 \cdot Energy_q(f) \cdot \left(E\left[Energy_q(g)\right]\right)^{\frac{1}{q}} + 4 \cdot \left(E\left[Energy_q(g)\right]\right)^{\frac{1}{q}}$$

- Jl embedding is as in the claim, thus $E\left[Energy_q(F)\right] \leq c \cdot \boldsymbol{OPT} + O\left(\frac{\sqrt{q}}{k}\right)$

■

Conclusions / Thank you slide

- We initiate theoretical study of measurement criteria widely used in practical applications
- We give theoretical bounds for these criteria, by showing that the JL random projection is, essentially, the optimal tool for dimensionality reduction
- Our bounds result in approximate algorithm for embedding any finite metric into k-dim space, with proven approximation ratio guarantees
- One of the central open questions arising from our work :
  close the gap for values of $q \leq \sqrt{k}$, and $q < 2$
- Thank you!