Embeddings

Andreas Galanis

May 13, 2010

Disclaimer

Most of the material presented here is in a straightforward manner adopted by a tutorial by Piotr Indyk at FOCS 2001 on *Algorithmic Aspects of Geometric Embeddings*.

1. Definition and Motivation

2. Embeddings of graph-induced metrics
   - into norms
   - into probabilistic trees

3. Embeddings of norms into norms
   - reduction of dimension

Definitions & Examples

Spaces (X, d_X)

- X set of points (finite or infinite).
- Metric distance function $d : X \times X \to \mathbb{R}_+$, i.e.

$$\begin{aligned} d(x,x) &= 0, && \forall x \in X. \\ d(x,y) &= d(y,x), && \forall x, y \in X. \\ d(x,y) &\leq d(x,z) + d(z,y), && \forall x, y, z \in X. \end{aligned}$$

Finite Metrics

- Denote $|X| = n$.
- Described by $\binom{n}{2}$ pairs of distances.
- Visualized by edge-weighted graphs.

Example:

$$X = \{a, b, c, d, e\}$$

	a	b	c	d	e
a	0	3	8	6	1
b		0	9	7	2
c			0	2	7
d				0	5
e					0

Infinite Metrics

We will mostly use $X = \mathbb{R}^k$ equipped with some Minkowski norm ℓ_p.

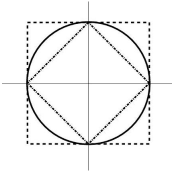
For $x \in \mathbb{R}^k$ its ℓ_p length is given by

$$\|x\|_p = \left(\sum_{i=1}^{k} |x_i|^p \right)^{1/p} \quad \text{for } 1 \leq p < \infty$$

For $x, y \in \mathbb{R}^k$, the ℓ_p-distance between them is $\|x - y\|_p$.

Some special cases:

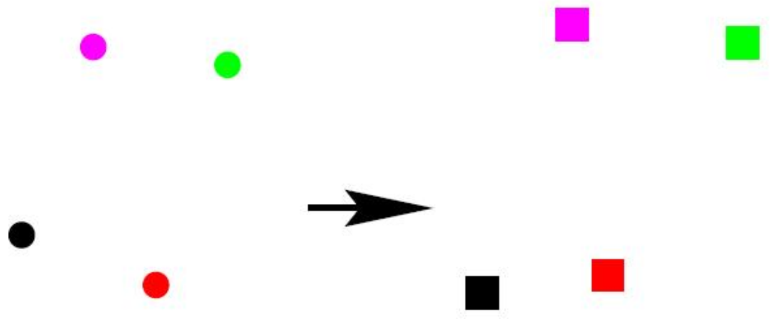
$$\begin{aligned} p = 1 &\quad \rightarrow \quad \text{Manhattan Distance} \\ p = 2 &\quad \rightarrow \quad \text{Euclidean Distance} \\ p = \infty &\quad \rightarrow \quad \|x\|_\infty = \max_{1 \leq i \leq k}\{|x_i|\} \end{aligned}$$



Unit balls

Embeddings

Given metrics (X, D) and (X', D') an *embedding* is a map $f : X \to X'$.



Embedding Finite Metrics to Weighted Graphs

A natural metric distance for weighted graphs is the length of the shortest path between vertices.

Conversely, a finite metric (X, D) can clearly be mapped into a weighted graph G such that:

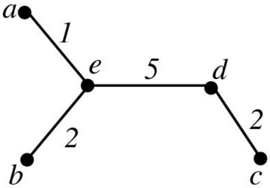
- Set X to be the vertices of the graph.
- Set the length of $\{i, j\}$ to $D(i, j)$.
- The shortest path metric in G clearly coincides with D.

In fact, we can drop edges in G as long as the shortest path metric is left invariant. The resulting minimal graph is called *critical graph*.

Example

$$X = \{a, b, c, d, e\}$$

	a	b	c	d	e
a	0	3	8	6	1
b		0	9	7	2
c			0	2	7
d				0	5
e					0

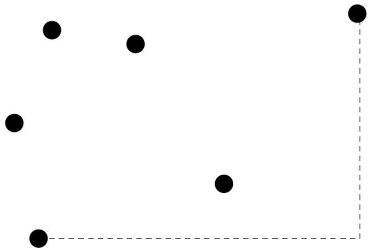


A Motivating Example

Why Embeddings?

Useful for reducing from "hard" to "easy" spaces.

- Given: a set P of n points in ℓ_1^d.
- Output the diameter of P, i.e. $\max_{p,q \in P} \|p - q\|_1$



- Easy to find in $O\left(dn^2\right)$.
- Can be solved in $O\left(nd2^d\right)$ by embedding ℓ_1^d in $\ell_\infty^{2^d}$.

Can we do better?

$\|p - q\|_1 = \sum_{i=1}^{d} \epsilon_i p_i - \sum_{i=1}^{d} \epsilon_i q_i$ for some choice of $\epsilon_i \in \{-1, 1\}$.

This suggests introducing 2^d vectors $y \in \{-1, 1\}^d$ and consider the inner products $y \cdot p$. Formally for every point $p \in P$:

1. Compute its inner products with each vector $y \in \{-1, 1\}^d$, i.e. $f_y(p) = y \cdot p$.
2. Concatenate these coordinates together, i.e. $f(p) = \bigoplus_{y \in \{-1,1\}^d} f_y(p)$.

$$\begin{aligned}
\max_{p,q \in P} \|p - q\|_1 &= \max_{p,q \in P} \left\{ \sum_{i=1}^{k} \epsilon_i p_i - \sum_{i=1}^{k} \epsilon_i q_i \right\} \\
\text{(not soo trivial)} &= \max_{p,q \in P} \max_{y \in \{-1,1\}^d} \{ f_y(p) - f_y(q) \} \\
&= \max_{p,q \in P} \|f(p) - f(q)\|_\infty
\end{aligned}$$

Thus it suffices to solve the problem in $\ell_\infty^{2^d}$.

Can we do better?

Solving the problem in $\ell_\infty^{2^d}$ is much easier:

$$\begin{aligned}\max_{x,y\in S}\|x-y\|_\infty &= \max_{x,y\in S}\max_{1\le i\le 2^d}|x_i-y_i| \\ &= \max_{1\le i\le 2^d}\max_{x,y\in S}|x_i-y_i|\end{aligned}$$

1. Solve the 1-dimensional problem in each of the 2^d coordinates.
2. Output the maximum over these values.

Properties of the Embedding

- Isometric.
- Linear.
- Deterministic.

A mapping $f : P_A \to P_B$:

- P_A: points from metric space with distance $D(\cdot, \cdot)$.
- P_B: points from some normed space, e.g. ℓ_2^d.
- For any $p, q \in P_A$

$$\frac{D(p,q)}{c} \leq \|f(p) - f(q)\| \leq D(p,q)$$

Parameter c is called "distortion".

Clearly $c \geq 1$. If $c = 1$ the embedding is called isometric.

Overview

- Embeddings of graph-induced metrics
    - into norms (Frechet's theorem, Bourgain's theorem, Matousek's theorem)
    - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
    - dimensionality reduction (Johnson-Lindenstrauss lemma)

Graph-induced Metrics into Norms

Let $G = (V, E)$. G induces the shortest path metric $D(\cdot, \cdot)$.

We will examine various embeddings of (V, D) into ℓ_p^d.

- General graphs $\rightarrow$ General Metrics.
- Special graphs (planar, trees, etc.) $\rightarrow$ Special Metrics.

Important parameters we seek to optimize:

- Dimension d.
- Distortion c.

General Finite Metric into ℓ_p^d

Bourgain (1985), Linial, London and Rabinovitch 1995

Any metric (X, D), and for any $p \geq 1$, can be embedded into ℓ_p^d with distortion $O(\log n)$ for $d = O\left(\log^2 n\right)$.

- Proof yields randomized algorithm with $O\left(n^2 \log^2 n\right)$ running time, can be derandomized.
- Suffices to prove the theorem for $p = 1$, the dimension ensures it extends easily for any $p \geq 1$.
- Matousek (1997) proved a stronger version of the theorem with distortion $O\left(\frac{\log n}{p}\right)$ for $1 \leq p < \log n$.

General Finite Metric into ℓ_∞^d

Matousek's Theorem (1996)

For any $b > 0$, any metric (X, D) can be embedded into ℓ_∞^d with distortion $c = 2b - 1$ for $d = O\left(bn^{1/b} \log n\right)$.

- It implies a weaker version of Bourgain's theorem for $b = O(\log n)$, with distortion $O\left(\log^2 n\right)$.
- Somewhat easier to derive, yet uses the same technique.

An Isometric Embedding into ℓ_∞^n

Frechet's Theorem

Any metric (X, D) can be embedded into ℓ_∞^n isometrically.

Let $X = \{p_1, p_2, \ldots, p_n\}$. Define $f(p) = \oplus_{1 \le i \le n} D(p, p_i)$.

We claim that $\|f(p_i) - f(p_j)\|_\infty = D(p_i, p_j)$.

- Shrinking secured by triangle inequality.

$$\|f(p_i) - f(p_j)\|_\infty = \max_{1 \le i \le n} |D(p, p_i) - D(p, p_j)| \le D(p_i, p_j).$$

- Expansion secured by the many dimensions.

$$\|f(p_i) - f(p_j)\|_\infty = \max_{1 \le i \le n} |D(p, p_i) - D(p, p_j)| \ge D(p_i, p_j).$$

In fact, the dimension can be reduced to $n - 1$.
For trees, the dimension can be reduced to $O(\log n)$.

Drawbacks of Isometric Embeddings

- Generally require high dimension (for example Frechet's theorem).
- Only ℓ_∞ has the universal property of Frechet's theorem.
  - C_4 cannot be embedded into ℓ_2 isometrically for any dimension!

Thus to obtain general results as Frechet's theorem, one needs to employ distortion.

Extensions

Instead of using points as "witnesses", use sets:

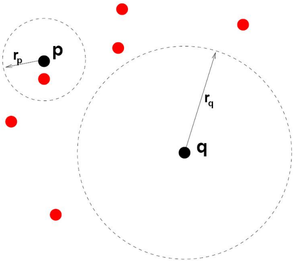
- $D(p, A) = \min_{a \in A} D(p, a)$.
- For carefully chosen sets $A_1, \ldots, A_{d'}$

$$f(p) = \oplus_{1 \le i \le d'} D(p, A_i)$$

Advantage: can achieve $o(n)$ dimensions.

Disadvantage: introduces distortion.

Ensuring Distortion



A_i=red dots

- $D(p, A_i) \leq r_p$
- $D(q, A_i) \geq r_q$

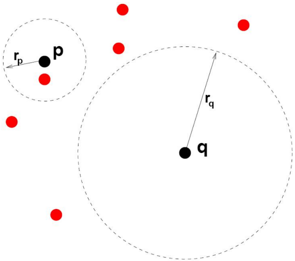
$|D(p, A_i) - D(q, A_i)| \geq r_q - r_p$

To show distortion c, we need $r_q - r_p \geq D(p, q)/c$.

Note $|D(p, A_i) - D(q, A_i)| \leq D(p, q)$ (using triangle inequality).

Find sets A_i with the above properties.

Constructing the sets A_i



Denote by B_p a ball centered at point p.

Two phases:

- Ensure existence of r_p, r_q such that the volume of B_p is not much smaller than the volume of B_q, an B_p, B_q disjoint (volume$\equiv$cardinality)
- Choose A_i's at random with proper density, so that with good probability it hits B_p and avoids B_q.

Ensuring the existence of balls B_p, B_q

Lemma: For each p, q there exists r such that

$$|B(p,r)| \geq \frac{|B(q, r + D(p,q)/c)|}{n^{1/b}}$$

or vice-versa, and the two balls are disjoint (recall that $c = 2b - 1$)

If we choose A_i by including each point to A_i with probability $\approx 1/|B(q, r + D(p,q)/c)|$, then with probability at least $\approx 1/n^{1/b}$:

- A_i hits $B(p, r)$.
- A_i avoids $B(p, r + D(p,q)/c)$.

To ensure success pick $n^{1/b} \log n$ subsets.
The problem is that we do not know neither r nor the cardinality of $B(q, r + D(p,q)/c)$.

Constructing the sets A_i

Generate A_i's using $\log n$ different probabilities $n^{-1/b}, n^{-2/b}, n^{-3/b}, \ldots$ to make sure we are OK for all densities.

To ensure success for each density, pick $n^{1/b} \log n$ subsets.

Total number of sets (which translates into dimension of embedding): $O(bn^{1/b} \log n)$.

Application on Cut Metrics

Consider a graph $G = (V, E)$ and a partition $S, \bar{S}$ of the set of its vertices.

A cut metric is such that:

$$d(x, y) = \begin{cases} 0 & \text{if both } x, y \in S \text{ or } x, y \in \bar{S} \\ 1 & \text{otherwise} \end{cases}$$

In sparsest cuts problems we wish to optimize over the cut metric. This induces a linear integer program. Instead:

- Relax the problem to an arbitrary metric taking values for $[0, 1]$
- Embed the resulting into ℓ_1 using Bourgain's Theorem.
- The ℓ_1 metric can be decomposed efficiently into a convex combination of cut metrics.
- Output the most suitable cut metric of these decompositions.

Bourgain's theorem can be used to obtain the best known bounds on such kind of problems, with an $O(\log n)$ factor in the approximation ratio.

Extensions

Volume respecting embeddings [Feige '98]:

- Stricter notion of embedding
- Ensures low distortion of k-dimensional "volumes":
  - Volume for a finite metric?
  - Largest among all of its contractions in $\mathbb{R}^{k-1}$.
  - Specializes to ordinary embedding for $k = 2$.

Overview

- Embeddings of graph-induced metrics
  - into norms (Frechet's theorem, Bourgain's theorem, Matousek's theorem)
  - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
  - dimensionality reduction (Johnson-Lindenstrauss lemma)

Probabilistic Metrics

Probabilistic metric is a convex combination of metrics:

- $T_1, T_2, \ldots, T_k$ are metrics, i.e. $T_i = (X, D_i)$.
- $\alpha_1, \ldots, \alpha_k > 0$: $\sum_i \alpha_i = 1$.
- The probabilistic metric $M = (X, \bar{D})$ is defined as:

$$\bar{D}(p, q) = \sum_i \alpha_i D_i(p, q)$$

Fix p, q and select T_i according to the weights α_i. Then

$$\mathbb{E}[D_i(p, q)] = \bar{D}(p, q)$$

Probabilistic Embeddings

Given

- a metric $M_Y = (Y, D)$
- a probabilistic metric $M_X = (X, \bar{D})$ defined by $T_i = (X, D_i), i = 1, \ldots, k$

a mapping $f : Y \to X$ is a probabilistic embedding of M_Y into M_X with distortion c if for any $p, q \in Y$:

- f expands by at most a factor of c on the average, i.e.

$$\bar{D}(f(p), f(q)) \leq cD(p, q)$$

- f never contracts, i.e, for each $i = 1, \ldots, k$

$$D_i(f(p), f(q)) \geq D(f(p), f(q))$$

Note the similarity with the general definition of embeddings (scale by $1/c$) but also the stronger second condition.

Embeddings into probabilistic trees

When each T_i is a tree (i.e. its critical graph is a tree).

Why

embed intro probabilistic trees?

Any cycle metric embeds into a tree metric with $\Omega(n)$ distortion. [Rabinovitch-Raz, Gupta'01]

Much better results for probabilistic trees (for any metric).

- AKPW'91: $2^{O(\sqrt{\log n \log\log n})}$ distortion.
- Bartal'96, Bartal'98: $O(\log^2 n)$ and $O(\log n \log\log n)$ distortion.
- FRT'04: $O(\log n)$ distortion. (Tight)

Many algorithmic applications, mostly on metrical task systems.

A weak version of Bartal's theorem

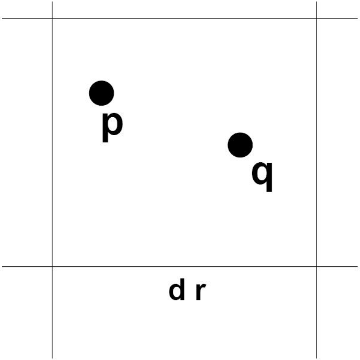
We will prove $O(\log^3 n \cdot \log \Delta)$ distortion, where Δ is the diameter of the original metric.

- Embed $M = (Y, D)$ into l_∞^d with distortion $O(\log n)$ and dimension $d = O(\log^2 n)$.
- Multiply final distortion by $O(\log n)$.
- Probabilistically partition the l_∞^d space into clusters of different diameters.
- Stitch the clusters together into a tree.

Probabilistic partitions

- ℓ-partition: any partition of Y into clusters of diameter $\leq \ell$.
- (r, ρ)-partition: a distribution over $r \cdot \rho$-partitions, such that for any $p, q \in Y$, the probability that p, q go to different clusters is at most $D(p, q)/r$.

In l_∞^d, (r, d)-partitions are easy to get by randomly shifting a grid of side $r \cdot d$.

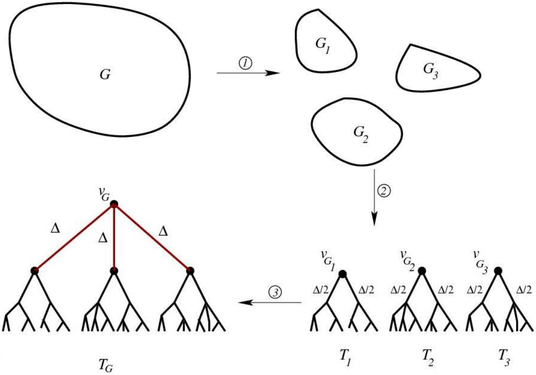


Probability of a cut between p and $q \leq d \cdot \dfrac{D(p,q)}{dr}$.

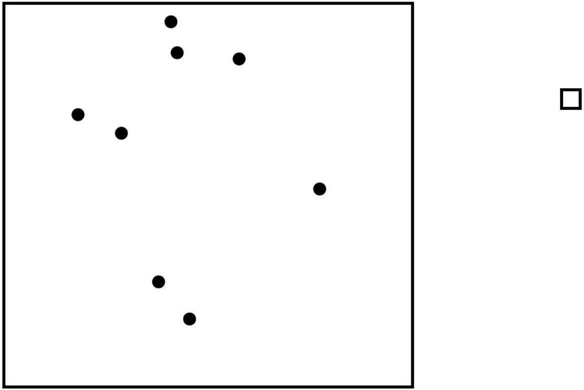
Probabilistic Tree Construction

Construction of a random tree. Initially $r = \Delta$.

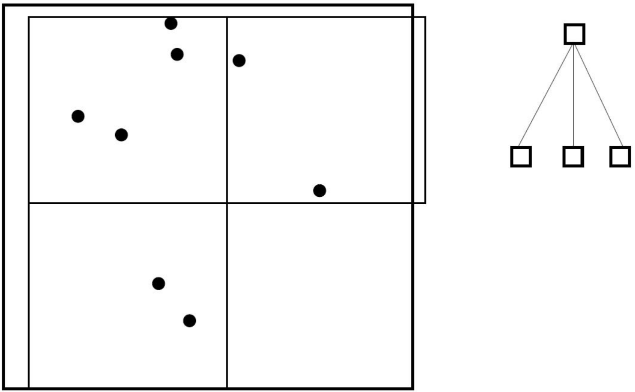
- Generate an $r \cdot d$-partition P from a (r, ρ)-partition.
- Within any cluster Y_i of P, generate a random tree T_i with root u_i using $r' \leftarrow r/2$
- Create new node u and connect u to u_i's using edges of length $r \cdot d/2$.



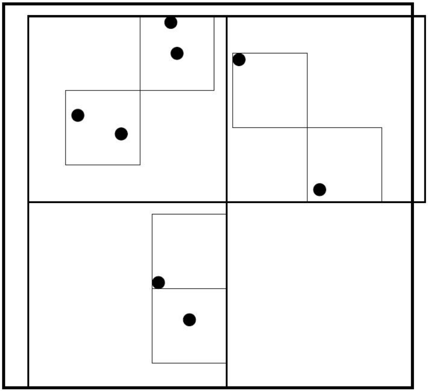
Probabilistic Tree Construction - Example

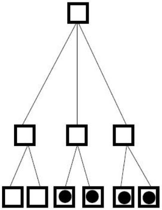


Probabilistic Tree Construction - Example

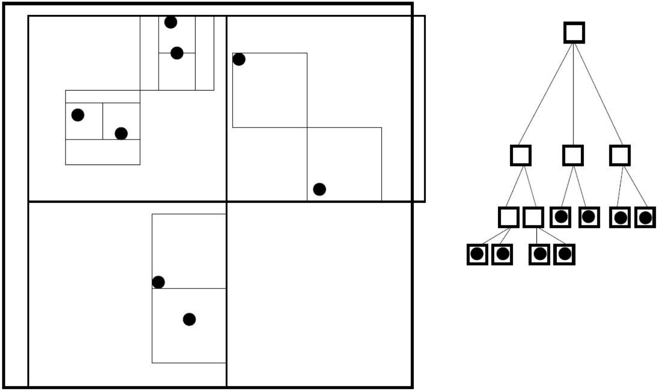


Probabilistic Tree Construction - Example





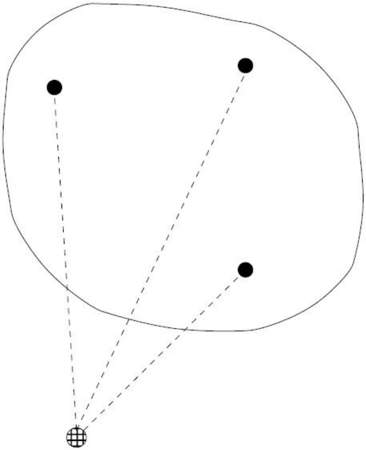
Probabilistic Tree Construction - Example



Contraction

No contraction, since:

- Consider any cluster Y of diameter $\leq rd$.
- Adding new node u with distance $rd/2$ to all points in Y cannot increase the distance.



Distortion

- One factor $\log n$ comes from embedding into l_∞^d.
- One factor comes from $\log \Delta$ levels in the tree.
- One factor $\log^2 n$ comes from d.

Distortion

Fix points $p, q \in Y$. The pair p, q:

- is separated by (Δ, d)-partition with probability $\frac{D(p,q)}{\Delta} \Rightarrow$ with probability at most $\frac{D(p,q)}{\Delta}$ level 1 contributes a total of tree distance $d \cdot \Delta$.
- is separated by $(\Delta/2, \rho)$-partition with probability $\frac{D(p,q)}{\Delta/2} \Rightarrow$ with probability at most $\frac{D(p,q)}{\Delta/2}$ level 1 contributes a total of tree distance $d \cdot \Delta/2$.
- ...

Expected distance:

- Per level: $\frac{D(p,q)}{\Delta/2^i} \cdot d \cdot \Delta/2^i = d \cdot D(p,q)$
- Summing over all levels: $O(d \log \Delta) \cdot D(p,q)$.

Applications on Online/Approximation Algorithms

Usually good guarantees for tree metrics. Thus for a metric M:

- Replace M by a random tree T.
- Solve the problem in T using the "good" algorithm.
- Interpret it as a solution in M

Competitive/Approximation ratio: Guarantee for trees $\times$ distortion of the embedding.

Overview

- Embeddings of graph-induced metrics
  - into norms (Frechet's theorem, Bourgain's theorem, Matousek's theorem)
  - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
  - dimensionality reduction (Johnson-Lindenstrauss lemma)

Reduction of Dimension in ℓ_2

Consider the space $\mathbb{R}^d$ with the Euclidean distance.

Is it possible

to embed a high dimensional pointset into a lower dimensional pointset with low distortion?

Not intuitively clear that this is possible.

Results are somewhat surprising at first glance.

JL-EMBEDDINGS

Johnson and Lindenstrauss '84: For every set P of n points in $\mathbb{R}^d$, then for every $\epsilon > 0$ and $k \geq k_0 = O(\epsilon^{-2} \log n)$, there exists $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in P$:

$$(1 - \epsilon) \|u - v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1 + \epsilon) \|u - v\|^2$$

- Original proof used heavy geometrical approximation tools.
- Frankl and Meahara '88: project onto k random orthonormal vectors.
- Indyk and Motwani '98: project onto k independent, spherically symmetric random vectors.
  - Pick each vector coordinate from a normal distribution independently.
  - The squared length of the embedded vector follows the chi-square distribution.
- Dasgupta and Gupta '99: same as the previous approach, but makes use of symmetry.

Analysis

- Pick k vectors, where each coordinate is taken from a normal distribution with mean 0 and variance 1.
- Project the n points onto a random k-dimensional hyperplane, i.e. for each point v in the original space, define $f(v)$ to be $\sqrt{\frac{d}{k}}v'$ where v' is the projection of v onto the hyperplane.

We need to analyze the distribution of the random variable $\frac{\|f(u) - f(v)\|^2}{\|u - v\|^2}$. Wlog $\|u - v\|^2 = 1$.

The distribution of $\|f(u) - f(v)\|^2$ is the same as that of a random unit vector projected onto a fixed k-dimensional hyperplane.

Thus, pick a random point on the unit d-dimensional sphere and project it onto the hyperplane defined by the first k coordinates.

Analysis

Picking a random point on the unit d-dimensional sphere:

- Generate vector $X = (X_1, \ldots, X_d)$, where each X_i follows $N(0, 1)$.
- Scale to obtain $Z = \frac{1}{\|X\|}(X_1, \ldots, X_d)$

Project onto the first k coordinates to obtain $Y = \frac{1}{\|X\|}(X_1, \ldots, X_k)$. Thus, it suffices to analyze

$$L = \|Y\|^2 = \frac{X_1^2 + \ldots + X_k^2}{X_1^2 + \ldots + X_d^2}$$

By symmetry $\mu = \mathbb{E}[L] = \frac{k}{d}$.

Analysis

We need to show concentration around the mean. Using Chernoff-type reasoning, it can be proved that

$$\Pr[L \leq (1-\epsilon)\mu] \leq \exp\left(\frac{-\epsilon^2 k}{4}\right)$$

$$\Pr[L \geq (1+\epsilon)\mu] \leq \exp\left(-\frac{k}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)\right)$$

Thus, for $k > \frac{4\ln n}{\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}}$, we have that

$$\Pr[|L-\mu| > \epsilon\mu] \leq 2\exp(-2\ln n) = \frac{2}{n^2}$$

Union bound for all $\binom{n}{2}$ pairs, yields that the embedding has the required property with probability $\geq \frac{1}{n}$.

Further Refinements

The value of k is tight as far as the previous analysis is concerned.

An interesting proof of the theorem was given by Achlioptas '04, where the vectors' coordinates are picked by the distributions:

$$r = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$

and

$$r = \sqrt{3} \times \begin{cases} +1 & \text{with probability } 1/6 \\ 0 & \text{with probability } 1/6 \\ -1 & \text{with probability } 1/6 \end{cases}$$

Analysis is more difficult since spherical symmetry is dropped.

For Further Reading

- Uriel Feige. *Approximating the bandwidth via volume respecting embeddings.*
- Anumam Gupta.*Algorithmic Applications of Metric Embeddings* (course).
- Piotr Indyk and Jiri Matousek.*Low distortion embeddings of finite metric spaces* (book chapter).