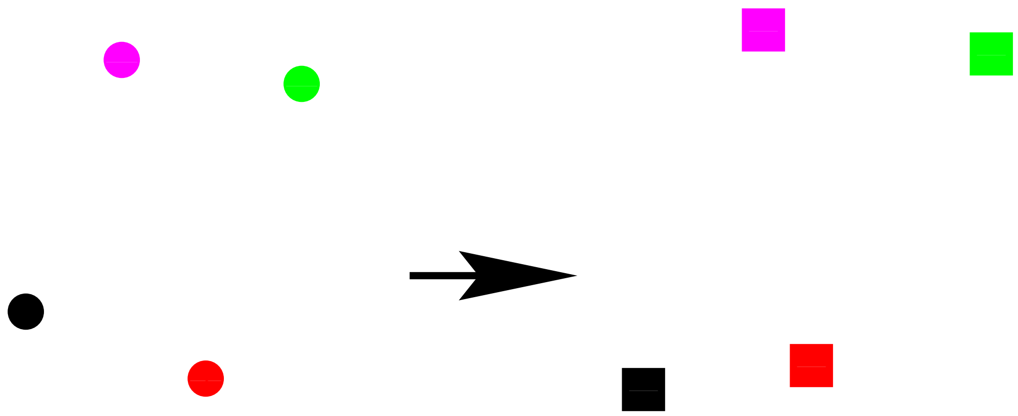
Algorithmic Applications of Low-distortion Geometric Embeddings

Piotr Indyk

MIT

Low-distortion geometric embeddings



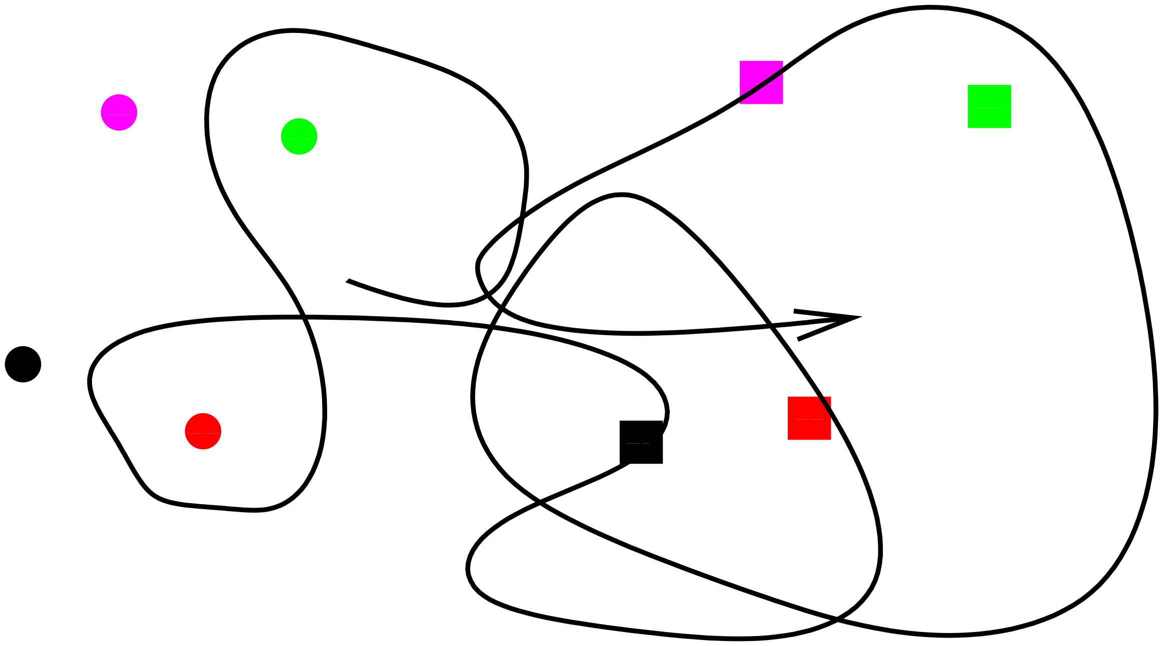
Formally: a mapping $f : P_A \rightarrow P_B$:

- 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 $p, q \in P_A$

$$1/c \cdot D(p,q) \leq \|f(p) - f(q)\| \leq D(p,q)$$

Parameter c is called "distortion".

Other embedding definitions possible

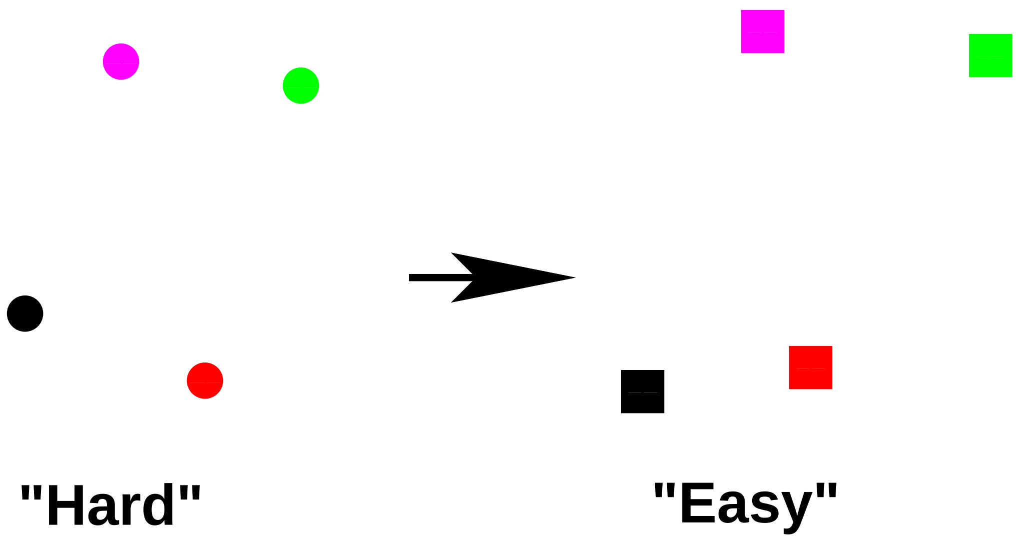


Overview of the remainder of the talk

- Motivation
  - General
  - Example: diameter in l_1^d
- Embeddings of finite metrics
  - into norms (Bourgain's theorem, Matousek's theorem, etc.)
  - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
  - dimensionality reduction (e.g., $l_2^{high} \rightarrow l_2^{small}$)
  - switching norms (e.g., $l_2 \rightarrow l_1$)
- Embeddings of special metrics into norms
  - string edit distance
  - Hausdorff metric

Why embeddings

- Reductions from “hard” to “easy” spaces:



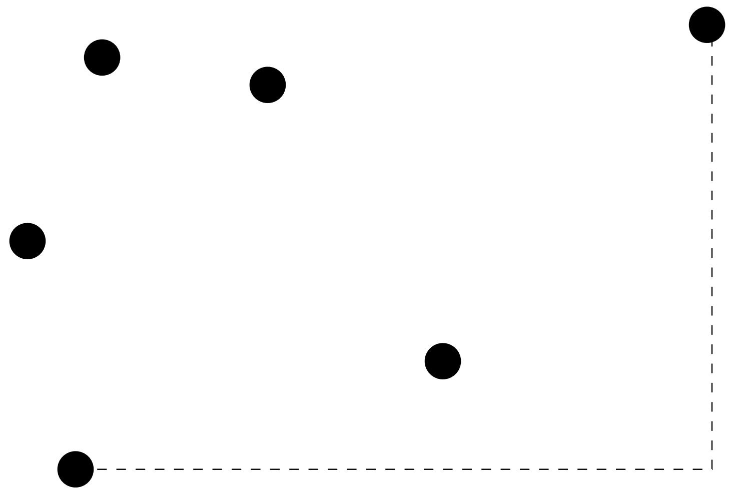
- Widely applicable

- Many tools available
  (combinatorics, functional analysis)

Example: diameter in l_1^d

- Given: a set P of n points in l_1^d
- Goal: the diameter of P, i.e.,

$$\max_{p,q \in P} \|p - q\|_1$$



Algorithms for diameter in l_1

- Easy: $O(dn^2)$ time
- Can we reduce the dependence on n (e.g., if d constant) ?

We will show $O(2^d n)$-time algorithm via:

- Embedding l_1^d into $l_\infty^{2^d}$
- Solving the problem in l_∞

Algorithm for diameter in $l_\infty^{d'}$

$$\max_{p,q\in P} \|p-q\|_\infty$$

$$=$$

$$\max_{p,q\in P} \max_{i=1\ldots d'} |p_i - q_i|$$

$$=$$

$$\max_{i=1\ldots d'} \left(\max_{p,q\in P} |p_i - q_i| \right)$$

$$=$$

$$\max_{i=1\ldots d'} \left(\max_{p\in P} p_i - \min_{q\in P} q_i \right)$$

Running time: $O(d'n)$.

Embedding l_1^d into $l_\infty^{2^d}$

The mapping f is defined as:

$$f(p) = < s_0 \cdot p,\ s_1 \cdot p,\ \ldots, s_{2^d-1} \cdot p >$$

where s_i is the ith vector in $\{-1,1\}^d$. Then

$$\|f(p) - f(q)\|_\infty = \|f(p-q)\|_\infty = \max_s |s \cdot (p-q)|$$

$$= \max_s |\sum_{i=1}^{d} s_i \cdot (p-q)_i| = |\sum_{i=1}^{d} \mathsf{sgn}((p-q)_i)(p-q)_i|$$

$$= \sum_{i=1}^{d} |(p-q)_i| = \|p-q\|_1$$

Running time: $O(d2^d n)$.

Properties of the embedding

- Isometry (distortion $c = 1$)
- Linear
- Oblivious: $f(p)$ does not depend on P
- Deterministic
- Explicit

Overview of the talk

- Motivation
  - General
  - Example: diameter in l_1^d
- **Embeddings of graph-induced metrics**
  - into norms (Bourgain's theorem, Matousek's theorem, etc.)
  - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
  - dimensionality reduction (Johnson-Lindenstrauss lemma, etc.)
  - switching norms
- Embeddings of special metrics into norms
  - string edit distance
  - Hausdorff metric

Embeddings of finite metrics into norms

Embeddings of $M = (X, D)$ into l_p^d

- X - finite set, $|X| = n$
- D - distance metric (symmetry, triangle inequality etc)
- $D(p, q)$ - shortest distance between p and q in some graph:
  - general graphs $\Rightarrow$ general metrics
  - planar graphs, trees etc $\Rightarrow$ more specialized metrics

General finite metric into norms

Bourgain's theorem (1985):

> Any $M = (X, D)$ can be embedded into l_2^d with distortion $O(\log n)$.

- d: originally exponential in n, can be reduced to $O(\log^2 n)$ [Linial-London-Rabinovitch'94]

- Proof yields randomized algorithm with $O(n^2 \log^2 n)$ running time, can be derandomized

Seminal result:

- Initiated the investigation of embedding finite metrics

- Introduced proof technique which works for other norms and graph classes

The l_∞ version

Matousek's theorem (1996):

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

- Implies $O(\log n)$-distortion embedding into $l_\infty^{\log^2 n}$ $\Rightarrow O(\log^2 n)$-distortion embedding into l_2
- Proof somewhat easier than Bourgain's proof
- Same technique

Proof: no-distortion case

Assume $c = 1$. Will show $d = n$ (Frechet, 1???).

Let $X = \{p_1, \ldots, p_n\}$. Consider a mapping f defined as:

$$f(p) = < D(p, p_1), \ldots, D(p, p_n) >$$

Need to show $|f(p) - f(q)|_\infty = D(p, q)$.

- f is a contraction, since for any $p_i \in X$

$$|D(p, p_i) - D(q, p_i)| \leq D(p, q)$$

$$\Rightarrow |f(p) - f(q)|_\infty = \max_{p_i} |D(p, p_i) - D(q, p_i)| \leq D(p, q)$$

- f does not "shrink" too much, since

$$|f(p) - f(q)|_\infty = \max_{p_i} |D(p, p_i) - D(q, p_i)|$$

$$\geq |D(p, p) - D(p, q)| = D(p, q)$$

Proof: general distortion

Modifications:

- "Witness" is a *set*, not a point, i.e.,
  - Define $D(p, A) = \min_{a \in A} D(p, a)$
  - Define

$$f(p) = < D(p, A_1), \ldots, D(p, A_d) >$$

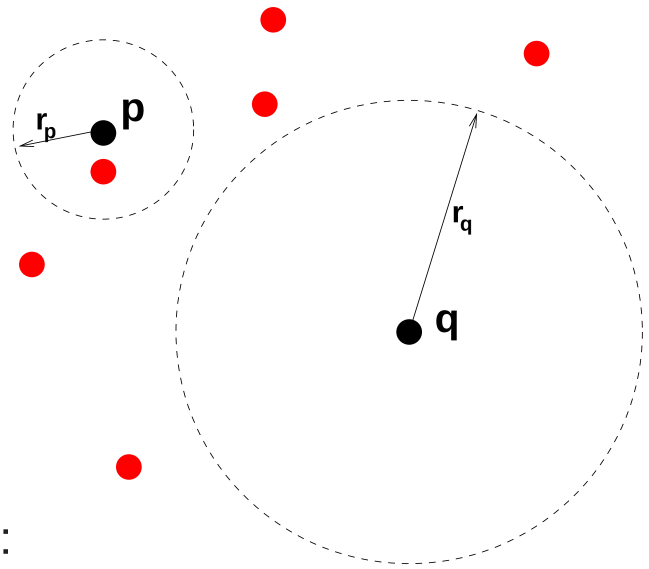
  for carefully chosen sets $A_i \subset X$

- Advantage: can achieve $d = o(n)$

- Drawback: "non-shrinking" only approximate, i.e., for any p, q there exists A_i such that

$$|D(p, A_i) - D(q, A_i)| \geq D(p, q)/c$$

Matousek's proof by picture



For each p, q:

1. There are $r_p, r_q > 0$, $r_q \geq r_p + D(p,q)/c$, and A_i, such that

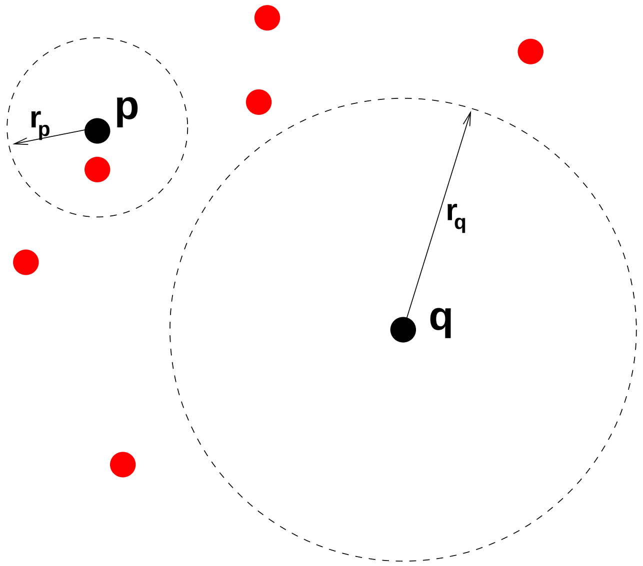
   - A_i hits the ball B_p of radius r_p around p
   - A_i avoids the ball B_q of radius r_q around q

   (or the same for p swapped with q).This implies

   $$|D(p, A_i) - D(q, A_i)| \geq D(p,q)/c, \text{ for some } A_i$$

2. $|D(p, A_i) - D(q, A_i)| \leq D(p,q)$ for all A_i
   (follows from triangle inequality)

Matousek's proof, ctd.



Need to construct the sets A_i (the red dots).

Main ideas:

1. Ensure existence of r_p, r_q such that the volume of B_p is not much smaller than the volume of B_q, and B_p, B_q disjoint (volume $\equiv$ cardinality)

2. Choose A_i's at random with proper density, so that with good probability it hits B_p and avoids B_q (prob. of including each point $\approx 1/$vol. of B_q)

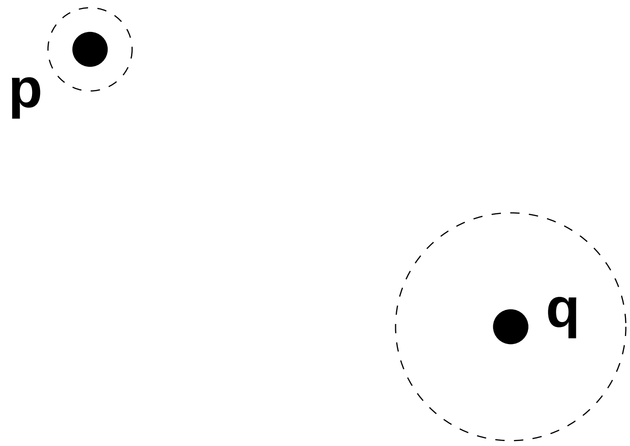
Main lemma

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

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

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

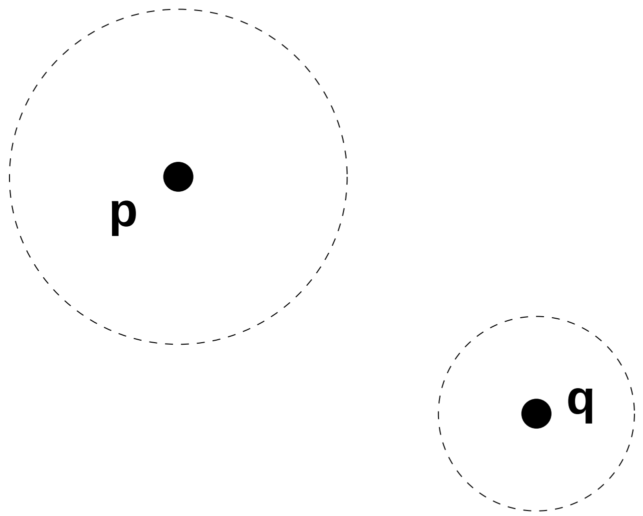
Proof: Start from $r = 0$. Check if $|B(p, 0)|$ not much smaller than $|B(q, D(p,q)/c)|$.



If so, we are done.

Main lemma: proof ctd.

Otherwise, swap the roles of p, q and take $r = D(p,q)/c$.



Check if $B(q,r)$ not much smaller than $B(p, r + D(p,q)/c)$. If so, we are done. Otherwise, repeat.

Observations:

- The process could take b steps until B_p, B_q overlap
- If the balls grew by $> n^{1/b}$ each time, they would have $> n$ volume at the end

Matousek's proof: the end

We know that there exists r such that

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

and the two balls are disjoint.

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 $\approx 1/n^{1/b}$:

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

Now:

- Generate $A_i s$ using $\log n$ different probabilities $1/2, 1/4, \ldots 1/n$ (to make sure we are OK for all densities)

- For each probability, generate $O(n^{1/b} \log n)$ sets A_i, to get a high probability of success

- Total number of sets: $O(n^{1/b} \log^2 n)$ (can be improved by a factor of $\log n/b$ using slightly different method)

Summing up

- Any metric can be embedded into l_∞^d with distortion $c = 2b - 1$, $d = O(bn^{1/b} \log n)$

- For $b = \log n$ we get $c = O(\log n)$, $d = O(\log^2 n)$ $\Rightarrow O(\log^2 n)$-distortion embedding into l_2

- Proof of Bourgain's theorem requires more "counting"

From	To	Distortion	Reference
any	l_2	$O(\log n)$	Bourgain'85
any	$l_\infty^{O(bn^{1/b}\log n)}$	$2b-1$	Matousek'96
expanders	$l_p, p = O(1)$	$\Omega(\log n)$	LLR'94
high girth graphs	any norm with dim $\Omega(n^{1/b})$	$2b-1$	Matousek'96 (Erdos conj.)
planar	l_2	$\Theta(\sqrt{\log n})$	Rao'99, Newman-Rabinovich'02
planar	$l_\infty^{\log^2 n}$	$O(1)$	
outerplanar	l_1	$O(1)$	GNRS'99
trees	l_1	1	folklore
trees	$l_\infty^{O(\log n)}$	1	LLR'94
trees	l_2	$\Theta(\sqrt{\log\log n})$	Matousek
(1,2)-metric with B 1's	$l_\infty^{O(B\log n)}$ (also l_p's)	1	Trevisan'97, I'00

Volume-respecting embeddings [Feige'98]

- Stricter notion of embedding
- Ensures low distortion of k-dimensional "volumes"
- Specializes to ordinary embedding for $k = 2$
- Proof uses Bourgain's technique in elaborate way (and implies Bourgain's theorem for $k = 2$)

Applications (of embeddings into norms)

- Approximation algorithms: Bourgain's theorem, volume-respecting embeddings
- Proximity-preserving labelling: Matousek's theorem
- Hardness results: (1,2)-metrics

App I: Approximation algorithms

Sparsest cut problem:

Given:

- graph $G = (V, E)$, cost $c : E \rightarrow \Re^+$
- k terminal pairs $\{s_i, t_i\}$, with demands $d(i)$

Goal: find $S \subset V$ minimizing

$$\rho(S) = \frac{\sum_{u \in S, v \in V-S} c(\{u, v\})}{\sum_{i : s_i \in S, t_i \in V-S} d(i)}$$

Sparsest cut: algorithm

- Long history, starting from [Leighton-Rao'88]
- Best so far: $O(\log k)$-approximation [Linial-London-Rabinovich'94, Aumann-Rabani'94]
- Method:
  - Solve linear relaxation of the problem - the solution forms a metric
  - Embed the metric into l_1
  - Solve the problem optimally assuming a metric induced by l_1
- Comments:
  - $O(\log k)$ comes from Bourgain's theorem
  - Easier metric $\Rightarrow$ better bounds (e.g., planar graphs)
  - Embedding does not provide a straightforward reduction

Applications of v. r. embeddings

- Min graph bandwidth: $\log^{O(1)} n$-approximation [Feige'98, Dunagan-Vempala'01]

- VLSI design problems [Vempala'98]

Again, embeddings do not provide straightforward reductions.

App II: Proximity-preserving labelling

Proximity-preserving labelling [Peleg'99]

- Given: a metric $M = (X, D)$, distortion c
- Goal: to find a *labelling* $f : X \to \{0,1\}^d$ such that
  - given $f(p), f(q)$ we can estimate $D(p,q)$ up to a factor of c
  - d as small as possible

Proximity-preserving labelling

Immediate application of low-distortion embeddings:

- Matousek's theorem gives best bound for general metrics
- Best isometric labelling scheme for trees also follows from embeddings
  (but not for constant tree-width graphs)

Implications in other direction [GPPR'01]:

- $\Omega(n^{1/2}/\log n)$ dimension lower bound for isometric embeddings of bounded degree graphs
- $\Omega(n^{1/3}/\log n)$ for bounded degree planar graphs

App III: Hardness

Necessity of double exponential dependence on d of PTAS's in l_p^d (e.g., for TSP) [Trevisan'97, I'00]

- Consider (1,2)-B metrics:
  - Distances 1 and 2,
  - At most B 1's per vertex, $B = O(1)$
- $(1+\epsilon)$-approximating TSP in such metrics is NP-hard [Papadimitriou-Yannakakis'87]
- But such metrics can be embedded into $l_p^{O(B \log n)}$
  - With very small distortion (and somewhat weaker def of embedding) for $p < \infty$ [Trevisan'97]
  - With no distortion for $p = \infty$ [I'00]
- Therefore, cannot have $2^{2^{o(d)}}$ time unless

$$\text{NP} \subset \text{DTIME}\left(2^{2^{o(\log n)}}\right) \subset \text{DTIME}\left(2^{o(n)}\right)$$

A digression

Embeddings used for all of the aforementioned applications:

- Approximation algorithms
- Proximity-preserving labelling
- Hardness (for l_∞)

are based on Bourgain's technique of "witness sets".

Overview of the talk

- Motivation
  - General
  - Example: diameter in l_1^d
- Embeddings of graph-induced metrics
  - into norms (Bourgain's theorem, Matousek's theorem, etc.)
  - **into probabilistic trees (Bartal's theorem)**
- Embeddings of norms into norms
  - dimensionality reduction (Johnson-Lindenstrauss lemma, etc.)
  - switching norms
- Embeddings of special metrics into norms
  - string edit distance
  - Hausdorff metric

Embeddings into probabilistic trees

Probabilistic metric is a convex combination of metrics, i.e.,

- if $T_1, \ldots, T_k$ are metrics, $T_i = (X, D_i)$
- and $\alpha_1 \ldots \alpha_n > 0$, $\sum_i \alpha_i = 1$
- then the prob. metric $M = (X, \overline{D})$ is defined by

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

If T_i chosen with probability α_i, then

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

Probabilistic embeddings

For

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

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

1. f expands by at most a factor of c *on the average*, i.e.,
$$\overline{D}(f(p), f(q)) \leq cD(p, q)$$
2. f *never* contracts, i.e,
$$\min_i D_i(f(p), f(q)) \geq D(p, q)$$

This is *more* than just an ordinary embedding of M_Y into M_X !

Why embed into probabilistic trees ?

Not possible to embed a cycle metric into a tree metric [Rabinovitch-Raz, Gupta'01] with $o(n)$ distortion.

Can do much better for probabilistic trees !
(for any metric)

- [AKPW'91]: $2^{O(\sqrt{\log n \log\log n})}$ distortion
- [Bartal'96] and [Bartal'98]:
  - $O(\log^2 n)$ and $O(\log n \log\log n)$ distortion
  - Simpler class of trees (Hierarchically Well-Separated Trees)
  - Many applications

Imply identical results for embeddings into l_1

Proof of weaker bound

We'll "show" $O(\log^3 n \cdot \log \Delta)$ distortion
(Δ - furthest/closest pair ratio)

Contains essential elements of [Bartal'96], with additional ideas.

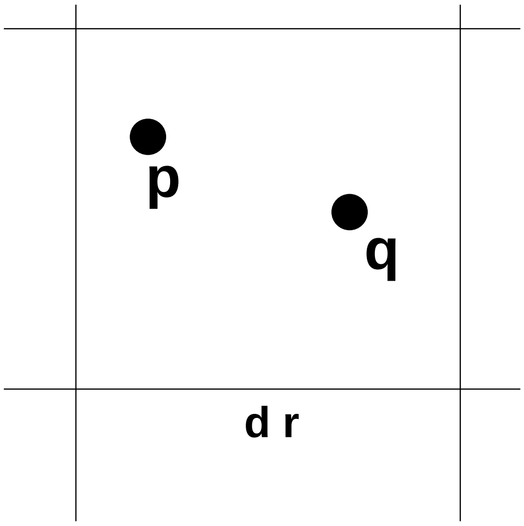
Proof:

- Embed $M = (Y, D)$ into l_∞^d with distortion $\log n$, $d = O(\log^2 n)$
- From now on, we assume M induced by l_∞, multiply final distortion by $\log n$
- Partition the l_∞^d space probabilistically into clusters of different diameters
- "Stitch" the clusters together into a tree

Probabilistic partitions

- l-partition: any partition of Y into clusters of diameter $\leq l$
- (r, ρ)-partition: a distribution over $r \cdot \rho$ partitions, such that for any $p, q \in Y$, the prob. 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 $r \cdot d$



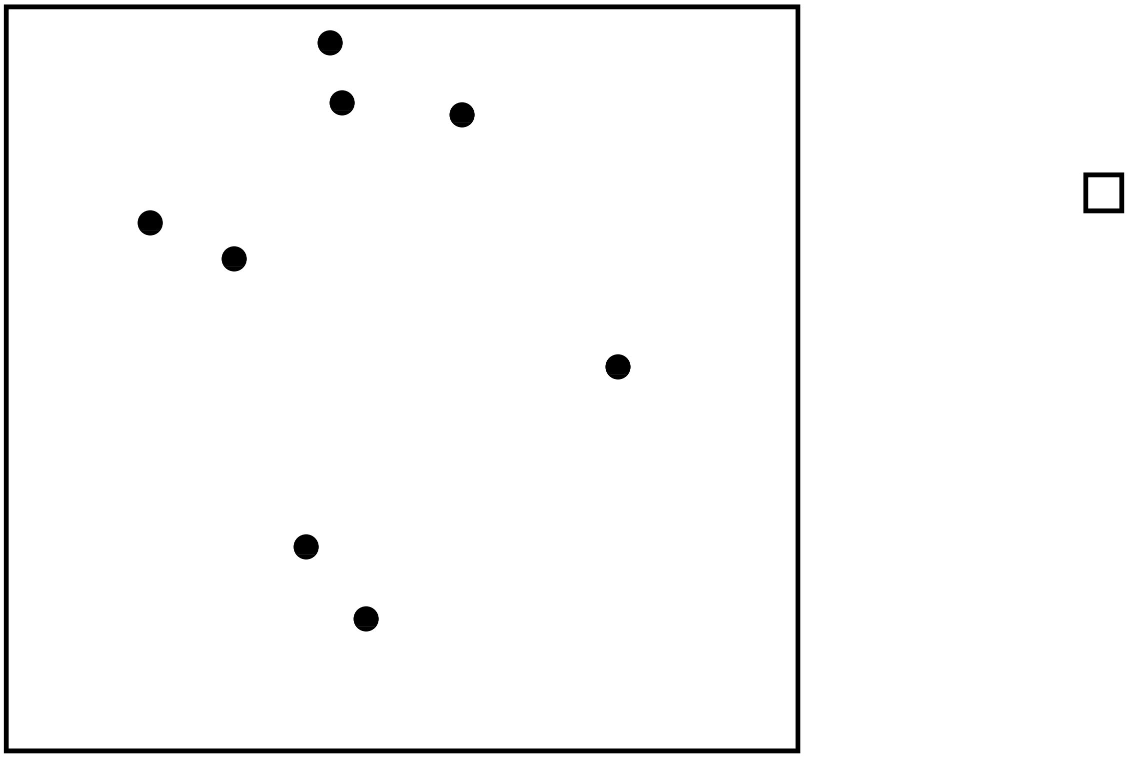
Probability of a cut $\leq d \cdot \frac{D(p,q)}{dr}$

Probabilistic tree construction

Recursive construction of a random tree. Initially $r = \Delta$.

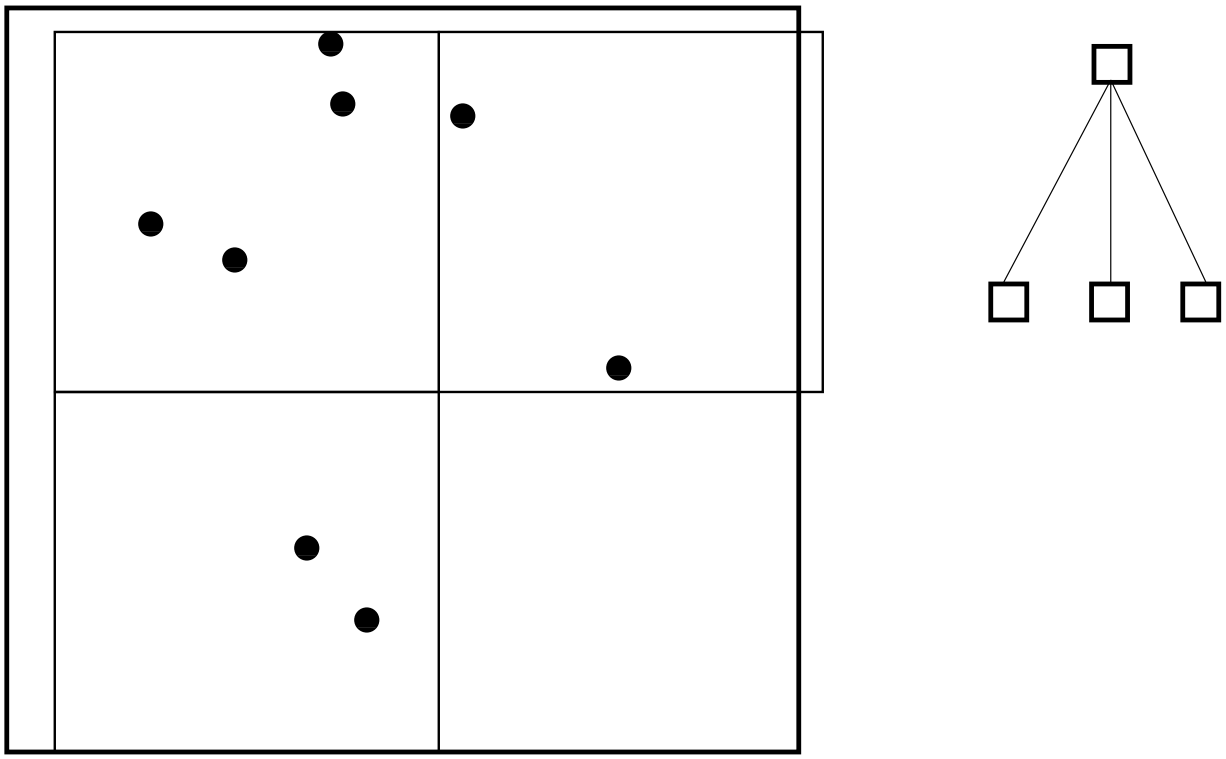
- Generate an $r \cdot \rho$-partition P from a (r, ρ)-partition
- Within any cluster Y_i of P, generate a random tree T_i with root u_i using $r/2$
- Create artificial node u and connect u to u_i's using edges of length $\rho \cdot r/2$

Construction: I



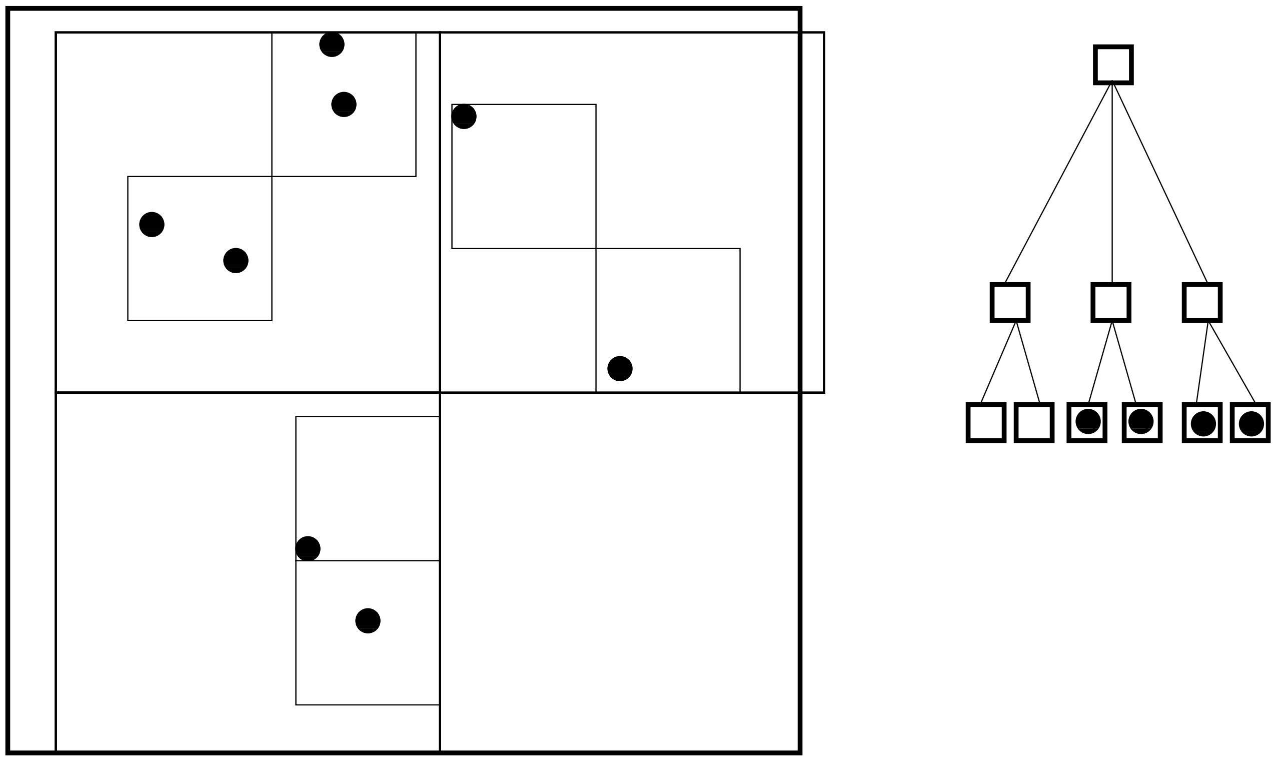
- Create a root
- We will create subtrees recursively

Construction: II



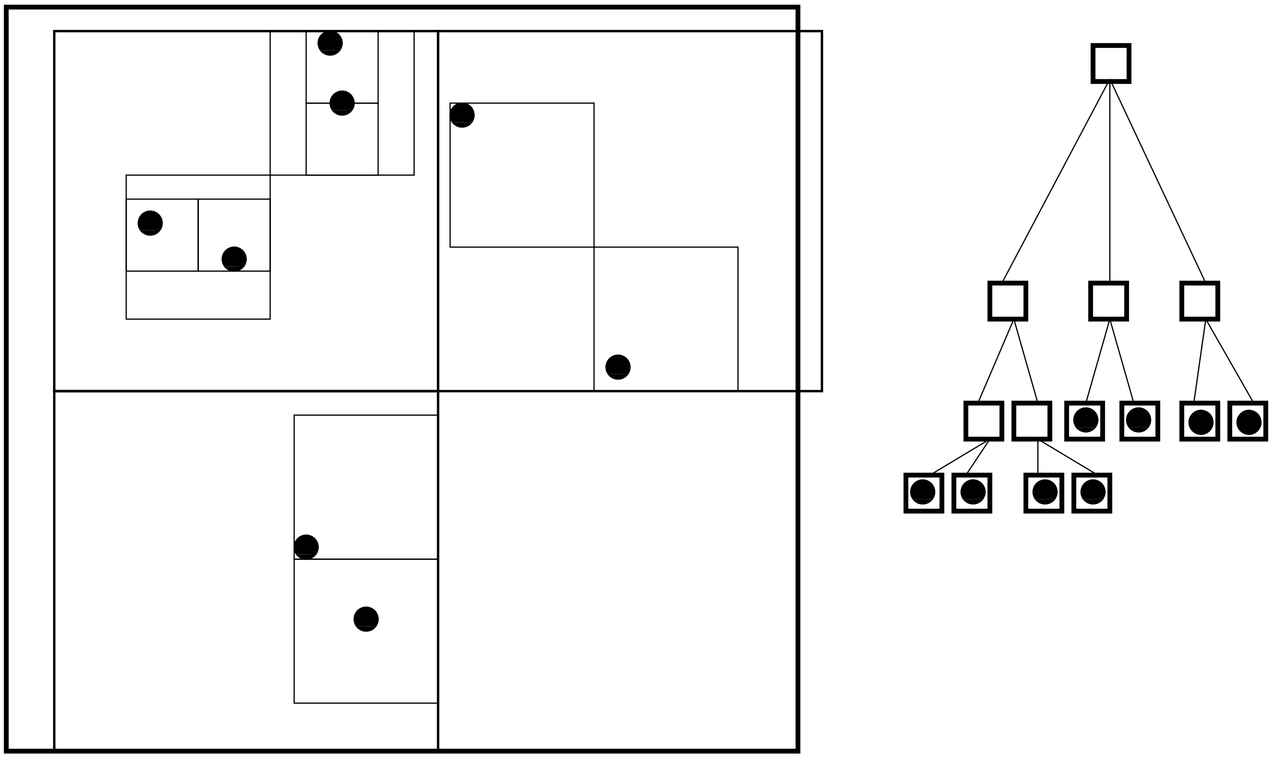
- Subdivide using a randomly shifted grid
- Create nodes for each cell
- Edge length proportional to the side of the grid cell
- Close points unlikely to be separated

Construction: III



- Further subdivide within each cell
- Stop when single points are reached

Construction: IV



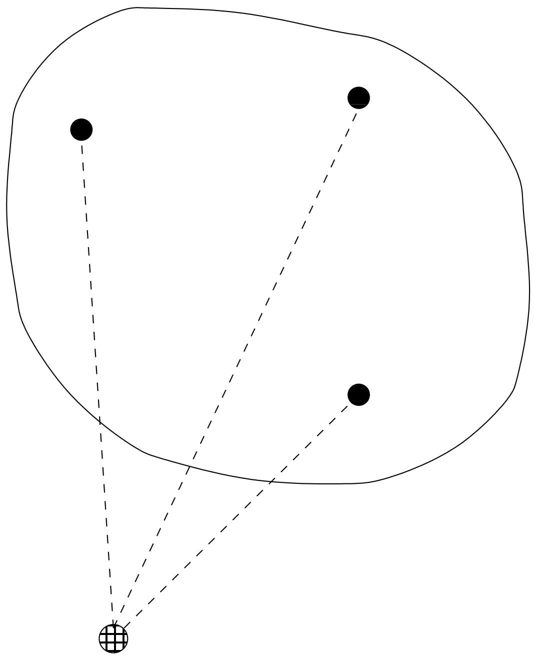
Distortion:

- One factor $\log n$ comes from embedding into l_∞
- One factor comes from $\log \Delta$ levels in the tree
- One factor $\log^2 n$ comes from ρ (ratio between probability of cutting and the edge length)

Non-contraction

No tree contracts the distances:

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



Distortion

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

- Is separated by (Δ, ρ)-partition with prob. $\frac{D(p,q)}{\Delta}$
  $\Rightarrow$ tree distance $\Delta \cdot \rho$

- Is separated by $(\Delta/2, \rho)$-partition with prob. $\frac{D(p,q)}{\Delta/2}$
  $\Rightarrow$ tree distance $\Delta/2 \cdot \rho$, etc...

Expected distance:

- $2^i r \cdot \rho \cdot \frac{D(p,q)}{2^i r} = \rho \cdot D(p,q)$ per level

- times $O(\log \Delta)$ levels

$= \underline{O(\rho \log \Delta)} \cdot D(p,q)$

Summing up

- Overall distortion: $O(\log^3 n \cdot \log \Delta)$
- Trees have special structure (HST):
  - On the way from the root to a leaf distances decrease exponentially
  - All distances from a node to its children are the same
- Can get rid of the additional nodes $\Rightarrow X = Y$

Summary of the prob. emb. into HSTs

From	Distortion	Reference
any	$O(\log n \log \log n)$	Bartal'98
high-girth	$\Omega(\log n)$	Bartal'96
planar	$O(\log n)$	GKR
l_2^d	$O(\sqrt{d} \log n)$	CCGGP'98

Applications (of embeddings into prob. trees)

Algorithms (approximate, on-line):

- Prob. embeddings provide fairly general reductions from problems over metrics to problems over trees
- Approximation algorithm for metric M:
  - Let A be an a-approximation algorithm for trees
  - Replace M by a random tree T
    $\Rightarrow OPT_T \leq c \cdot OPT_M$
  - Use A on T to produce a solution for T with cost $\leq a \cdot OPT_T \leq a \cdot c \cdot OPT_M$
  - Interpret it as a solution for M
  - Final cost $\leq a \cdot c \cdot OPT_M$
- Similar approach works for on-line problems
- The structure of HST makes the task even easier

Applications: on-line algorithms

Metrical task systems [Borodin,Linial,Saks'87]:

- Defined by a metric $M = (X, D)$, initial *server* position $p \in X$
- Input: a sequence of tasks $\tau = \tau_1, \tau_2, \ldots$, $\tau_i : X \to [0, \infty)$
- Given next task τ_i, the algorithm:
  - Moves the server from its current position x to a new position y
  - Serves the task from y
  - Incurred cost: $D(x, y) + \tau(y)$
- Want: to design an algorithm A with small competitive ratio, i.e.,

$$\max_{\tau} \frac{\text{Cost incurred by } A \text{ on } \tau}{\text{Optimal cost of serving } \tau}$$

Prob. embeddings for MTS

- We have seen prob. embedding of $M = (X, D)$ into $(X, \overline{D})$, where $(X, \overline{D})$ is a convex combination of HSTs

- Can use it to reduce the problem over general metrics to problem over HSTs:
  - Let A be a b-competitive algorithm for HST
  - Choose a random HST T
  - Feed all tasks to A
  - Interpret all server moves of A as moves in M

- Cost estimations:
  - Let OPT be optimal server trajectory in M with cost C
  - It corresponds to a server trajectory in T with expected cost $\leq c \cdot C$, where c is the distortion
  - A will find a solution S for T with cost $\leq b \cdot c \cdot C$
  - Interpreting S as a solution for M only decreases the cost

Applications of prob. embeddings

- For "metric" problems, a b-competitive algorithm for HSTs implies a (randomized) $O(b \log^{O(1)} n)$-competitive algorithm for general metric:
  - $O(\log^{O(1)} n)$-competitive algorithm for metrical task systems [BBBT'98,FM'00]
  - distributed problems [Bartal'98]
- Same holds for approximation algorithms:
  - "Buy-at-bulk" network design [Azar-Awerbuch'97]
  - Group Steiner problem
  - ...(≈ 10 problems)

Overview of the talk

- Motivation
  - General
  - Example: diameter in l_1^d
- Embeddings of graph-induced metrics
  - into norms (Bourgain's theorem, Matousek's theorem, etc.)
  - into probabilistic trees (Bartal's theorem)
- **Embeddings of norms into norms**
  - dimensionality reduction (Johnson-Lindenstrauss lemma, etc.)
  - switching norms
- Embeddings of special metrics into norms
  - string edit distance
  - Hausdorff metric

Embeddings of norms into norms

Different from finite metrics:

- Embeddings of *infinite* spaces
- Advantage: we do not have to know all points in advance
- Drawback: sometimes guarantees only randomized

Randomized embeddings

For metrics $M = (X, D), M' = (X', D')$, a distribution $\mathcal{F}$ over mappings $f : X \to X'$ is a randomized embedding with

- distortion c
- contraction probability P_{con}
- expansion probability P_{exp}

if for any $p, q \in X$ we have

- $D'(f(p), f(q)) < 1/c \cdot D(p, q)$ with prob. $\leq P_{con}$
- $D'(f(p), f(q)) > D(p, q)$ with prob. $\leq P_{exp}$

$P = P_{con} + P_{exp}$ is called *failure probability*

Dimensionality reduction in l_2

Johnson-Lindenstrauss (1984):

> There is a randomized embedding from l_2^d into $l_2^{d'}$ with distortion $1+\epsilon$ and failure probability $e^{-\Omega(d'/\epsilon^2)}$.

Corollary: For any set $P \subset l_2^d$ there exists an embedding of (P, l_2) into $l_2^{d'}$ with distortion $1+\epsilon$, where $d' = \frac{const}{\epsilon^2} \cdot \ln|P|$.

($const \approx 4$ for small enough $\epsilon > 0$)

Proof

- Several proofs known [JL'84,FM'88,IM'98,DG'99,AV'99]

- All of them proceed by showing:

  Take any $u \in \Re^d$, $\|u\|_2 = 1$.
  Let $A_1, \ldots A_{d'}$ be "random" vectors from $\Re^d$, and let $A = [A_1 \ldots A_{d'}]^T$. Then $\|Au\|_2$ is sharply concentrated around its mean (equal to 1).

- Linearity of A implies that for $p, q \in l_2^d$, we have

$$\|Ap - Aq\|_2 = \|A(p-q)\|_2 = \|p-q\|_2 \cdot \|Au\|_2 \approx \|p-q\|_2$$

where $u = (p-q)/\|p-q\|_2$.

Proof (sketch)

We show a proof when all entries in A chosen from Gaussian distribution $N(0,1)$ [I-Motwani'98]

- Sum of independent random variables from Gaussian distribution has Gaussian distribution
  $\Rightarrow$ each $A_i u$ has Gaussian distribution

- The variance of a sum is a sum of variances
  $\Rightarrow$ the variance of each $A_i u$ is $\sum_j u_j^2 = 1$
  $\Rightarrow$ each $A_i u$ is indep. chosen from $N(0,1)$

- $\|Au\|_2^2$ is a sum of squares of independent Gaussians
  - sum of squares of two Gaussians has exponential distribution
  - sum of squares of many Gaussian has chi-square distribution
  - the distributions well understood
  - "Plug and Play"

Summary of the results

- Distortion: $1+\epsilon$
- Prob. of contraction: P_{con}
- Prob. of expansion: P_{exp}
- Failure probability $P = P_{con} + P_{exp}$

Norm	Dimension	Reference
l_2	$O(\log(1/P)/\epsilon^2)$	JL'84
l_2	$\Omega(1/\log(1/\epsilon) \cdot \log(1/P)/\epsilon^2)$	A+C+M
l_1	$(\log(1/P_{con}) + 1/P_{exp})^{O(1/\epsilon)}$	I'00
Hamming (dist. range)	$O(\log(1/P)/\epsilon^2)$	KOR'98 I'00

Techniques used

- l_2 upper bound: random projection on a plane spanned by set of random vectors
  - chosen i.i.d. from d-dim Gaussian distribution (can be efficiently derandomized [EIO'02])
  - chosen i.i.d. from uniform dist. over a sphere
  - forced to be orthonormal (Haar measure) [JL,FM]
  - chosen i.i.d. from $\{-1,1\}^d$ or $\{-1,0,1\}^d$ [Achlioptas'01]

  Can be derandomized using [Shivakumar'02]

- l_2 lower bound: upper bound on "almost orthogonal" vectors in $\Re^d$ [Alon, Charikar, Matousek]

- l_1 upper bound: 1-stable distributions, i.e., generate A such that $\|Ax\|_1$ estimates $\|x\|_1$

- Hamming metric: random linear mapping over GF(2)

Application of dimensionality reduction

- “Straightforward” applications
- Faster embedding computation
- Continuous (clustering) problems
- Sublinear-storage computation
- Miscellaneous:
  - learning robust concepts [Arriaga-Vempala’99]
  - deterministic approximation algorithms using semidefinite programming [Engebretsen-I-O’Donnell’02, Shivakumar’02]

App I: Straightforward applications

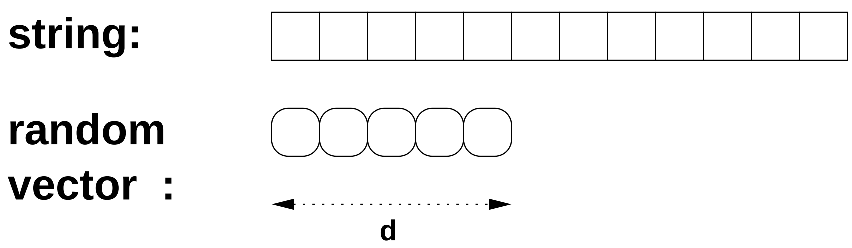
Running time:

$T(n,d) \Rightarrow T(n, \log n) + d \log n \cdot$ (# points to embed)

- Linear improvement: closest pair, nearest neighbor, diameter, MST etc.
  - time: $O(dn^2) \Rightarrow O(\log n \cdot n^2) + O(dn \log n)$
- Exponential improvement: nearest neighbor [Kushilevitz-Ostrovsky-Rabani'98, I-Motwani'98]
  - space: $n2^{O(d)} \Rightarrow n^{O(1)}$
  - query: $(d + \log n)^{O(1)} \Rightarrow O(d \log n + \log^{O(1)} n)$

App II: Faster embedding computation

- Computing embedding in $o(dn)$ time
- Feasible if the pointset defined implicitly, e.g., as a set of all d-substrings of a given string
- A *substring difference* problem: preprocess the data to estimate (quickly) the distance between two given d-substrings [I-Koudas-Muthukrishnan'00]
  – dim. reduction gives $O(n \log n)$ space and $O(\log n)$ query time
  ... but $\Theta(dn \log n)$ preprocessing time
  – embedding linear $\Rightarrow$ can use FFT to get $O(n \log d \log n)$ preprocessing time



App II: Faster embedding computation, ctd.

- Other string problems: variable d, string nearest neighbor problem [I-Koudas-Muthukrishnan'00]

- Line crossing metric [Har-Peled-I'00]

App III: Continuous (clustering) problems

- Generic problem:
  - Given: n points in l_p^d
  - Find: k *centers* in $\Re^d$ to minimize the total distance between the points and their nearest centers

    (total distance $\in \{$max dist., sum of dist.,$\ldots\}$)

- Simple dimensionality reduction does not work! (solution in the reduced space could be bogus)

- Idea [Dasgupta'99]:
  - Reduce the dimension
  - Identify (or guess) the clusters (not centers!) in the low-dimensional space
  - For each cluster, find its center in original space

- Works for learning mixtures of Gaussians [D'99], k-median for small k [OR'00], k-center

Low-storage computation

- Dimensionality reduction reduces *space* as well
- Prototypical example: vector maintenance
  - Data structure maintaining $x \in \Re^d$
    (x_i - counter for element i)
  - Enables increments/decrements of coordinates
  - Reports estimation of $\|x\|_p$
- Applications:
  - $p = 0$: # non-zero positions (distinct elements)
  - $p = 2$: self-join size

Norm maintenance: results

$(1+\epsilon)$-approximation in $(\log n + 1/\epsilon)^{O(1)}$ space:

- $p = 0$ (but $x \geq 0$): Flajolet-Martin'85
- $p = 2$: Alon-Matias-Szegedy'96
  (also any integer p with sublinear storage)
- $p \in [0, 2]$: I'00, Cormode-Muthukrishnan'01
  (earlier FKSV'99,FS'00)

Norm maintenance: approach

- Maintain low-dimensional Ax to represent x
- Reduce the amount of randomness used in A
- Implementation:
  - [AMS'96]:
    - 4-wise independent entries of A
    - Use median (not sum) to estimate the norm
  - [I'00]:
    - Use Nisan's generator to generate A
    - Can "simulate" JL lemma
    - Works for any $p \in [0, 2]$ via p-stable distributions

Other low-storage results

- Maintaining string properties [CM'01]
- Norm maintenance in fixed window [DGIM'02]
- Maintaining approximations of a vector (wavelet [GKMS'01], piecewise-linear [GGIKMS'01])
- ...

Overview of the talk

- Motivation
  - General
  - Example: diameter in l_1^d
- Embeddings of graph-induced metrics
  - into norms (Bourgain's theorem, Matousek's theorem, etc.)
  - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
  - dimensionality reduction (Johnson-Lindenstrauss lemma, etc.)
  - **switching norms**
- Embeddings of special metrics into norms
  - string edit distance
  - Hausdorff metric

Switching norms

- We have seen one already ($l_1 \rightarrow l_\infty$)
- Mostly ordinary embeddings, at last!
  (although often constructed using random mappings)
- Switch from "hard" to "easy" norms (l_1 or l_∞)
- All constructed using linear mappings
- Topic extensively investigated in functional analysis

Embeddings

Embeddings from l_p^d into $l_1^{d'}$

From	Dist.	d'	Reference	
$p = 2$	$1+\epsilon$	$O(d \log(1/\epsilon)/\epsilon^2)$	FLM'77	ala JL
	$\sqrt{2}$	$O(d^2)$	Berger'97	explicit
	$1+\epsilon$	$d^{O(\log d)}$	I'00	explicit
$p \in [1,2]$	$1+\epsilon$	$O(d \log(1/\epsilon)/\epsilon^2)$	JS'82	

Embeddings from l_p^d into $l_\infty^{d'}$

From	Dist.	d'	Reference
$p = 1$	1	2^{d-1}	folklore
polyhedral norm	1	$F/2$ ($F = \#$ faces)	folklore
any norm	$1+\epsilon$	$O(1/\epsilon)^{d/2}$ (Dudley's theorem)	folklore
$p = 2$	$1+\epsilon$	$(\log(1/P_{con}) + 1/P_{exp})^{O(1/\epsilon)}$	I'01

Applications of norm switching

- Embeddings into l_1 norm
  - $l_2 \rightarrow l_1 \rightarrow$ Hamming: approx. nearest neighbor algorithms [Kushilevitz-Ostrovsky-Rabani'98, I-Motwani'98]
  - same route: k-median algorithm [Ostrovsky-Rabani'00]

- Embeddings into l_∞ norm
  - Diameter/furthest neighbor in l_1, l_2
  - Nearest neighbor in product of l_2 norms [I'01]

Overview of the talk

- Embeddings of graph-induced metrics
  - into norms (Bourgain's theorem, Matousek's theorem, etc.)
  - into probabilistic trees (Bartal's theorem)
- Embeddings of norms into norms
  - dimensionality reduction (Johnson-Lindenstrauss lemma, etc.)
  - switching norms
- **Embeddings of special metrics into norms**
  - string edit distance
  - Hausdorff metric

Special metrics

- Hausdorff metric: for any two *sets* $A, B \subset X$ in a metric $M = (X, D)$, define

$$\overrightarrow{D_H}(A, B) = \max_{a \in A} \min_{b \in B} D(a, b)$$

$$D_H(A, B) = \max(\overrightarrow{D_H}(A, B), \overrightarrow{D_H}(B, A))$$

  Applications: vision, pattern recognition
  $(M = l_2^2, l_2^3)$

- Levenstein metric: $D_L(s, s')$ = minimum number of insertions/deletions/substitutions/etc. needed to transform s into s'

  Applications: computational biology, etc.

Special metrics

- Would like to solve problems (e.g., nearest neighbor, clustering) over D_H, D_L

- However, these metrics are more complex than normed spaces
  - D_H "contains" l_∞
  - D_L "contains" Hamming metric

- Thus, would like to embed them into proper normed spaces

- Additional benefit: if embedding is fast, can get fast approximate algorithm for computing $D(\cdot, \cdot)$

Embeddings of special metrics

From	To	Dist.	Dim.	Ref
D_H over (X, D)	l_∞	1	$\lvert X \rvert$	FI'99
D_H over l_p^d (s-subsets)	l_∞	$1+\epsilon$	$s^2/\epsilon^{O(d)}$	FI'99
D_L with block moves	Hamm.	$\approx \log d$		CPSC'00, MS'00,CM'01

Other metrics:

- Permutation distances [Cormode-Muthukrishnan-Sahinalp'01]

Conclusions

- We have seen lots of embeddings!
- But also main techniques used:
  - Finite metrics: “witness sets”
  - Normed spaces: random linear mappings
  - Probabilistic trees: stitching prob. partitions into trees
- Tools mostly taken from combinatorics and functional analysis

Open problems

- General open problems:
  - More embeddings
  - More applications of embeddings

- Specific problems:
  - Planar graph metrics into l_1
  - $O(\log n)$ distortion for embedding metrics into probabilistic trees
  - Dimensionality reduction for l_1
  - Embeddings of Levenstein metric