The Art of Metric Embeddings

— A technique oriented approach

Qin Zhang

HKUST

August 17, 2007

Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

Metric space

A metric on a set of points X is a distance function $d : X \times X \rightarrow \mathbb{R}$. For all points x, y, z in X, this function is required to satisfy the following conditions:

1. Non-negativity: $d(x, y) \geq 0$.
2. Identity of indiscernibles: $d(x, y) = 0$ if and only if $x = y$.
3. Symmetry: $d(x, y) = d(y, x)$.
4. The triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$.

The methodology

- Embedding

Original space (X, d_X) $\xrightarrow{f}$ Host space (Y, d_Y)

The methodology

- Embedding

  Original space (X, d_X) $\xrightarrow{f}$ Host space (Y, d_Y)

- e.g. furthest pair problem

  l_1^k: $O(kn^2)$
  by brute force

The methodology

- Embedding

Original space (X, d_X) $\xrightarrow{f}$ Host space (Y, d_Y)

- e.g. furthest pair problem

l_1^k: $O(kn^2)$
by brute force

l_∞^d: $O(nd)$
$\max\{$furthest pair along i-th coordinate : $1 \leq i \leq d\}$

The methodology

- Embedding

  Original space (X, d_X) $\xrightarrow{f}$ Host space (Y, d_Y)

  - e.g. furthest pair problem

    l_1^k: $O(kn^2)$
    by brute force

    isometric f $\longrightarrow$ $d = 2^k$

    l_∞^d: $O(nd)$
    $\max\{$furthest pair along i-th coordinate : $1 \leq i \leq d\}$

The methodology

- Embedding

  Original space (X, d_X) $\xrightarrow{f}$ Host space (Y, d_Y)

  - e.g. furthest pair problem

    l_1^k: $O(kn^2)$ by brute force $\xrightarrow[d = 2^k]{\text{isometric } f}$ l_∞^d: $O(nd)$ $\max\{$furthest pair along i-th coordinate : $1 \leq i \leq d\}$

- But ... Distortion is inevitable sometimes!

  - What is the best host metric space?
  - How small can the distortion and dimension of the host space be?

Preliminaries

- Distortion

X: a metric space with distance function d_X.
Y: a metric space with distance function d_Y.

A mapping $f : X \to Y$ is called L-Lipschitz if

$$\forall x, y \in X, \quad d_Y(f(x), f(y)) \leq L \cdot d_X(x, y).$$

We say f is an *embedding with distortion D* $(D \geq 1, D \in \mathbb{R})$ if there exists $r > 0$ such that

$$\forall x, y \in X, \quad r \cdot d_X(x, y) \leq d_Y(f(x), f(y)) \leq D \cdot r \cdot d_X(x, y)$$

Preliminaries (cont.)

- **Scale**

  Scale is just a measure of the length.
  We say (x, y) is *an edge at scale δ* if $d(x, y) \in [\delta, 2\delta]$.

- **Domination**

  Metric space X over point set V *dominates* metric space Y over V if $\forall u, v \in V, \quad d_X(u, v) \geq d_Y(u, v)$.

- **Graph representation**

  We can visualize a finite metric (X, d) as a weighted (undirected) graph $G(d)$: let the points in X be the vertices of the graph, and $d(x, y)$ be the weight of the edge (x, y).

  The *shortest path distance* in $G(d)$ is equal to the distance d in original metric.

Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

Three major themes

- Embed general metric into norms
  - Bourgain's theorem: general $\longrightarrow l_2$, distortion $O(\log n)$
  - Extension: volume respecting embedding, general $\longrightarrow$ $\mathtt{NEG}\ldots$

Three major themes

- Embed general metric into norms
  - Bourgain's theorem: general $\longrightarrow l_2$, distortion $O(\log n)$
  - Extension: volume respecting embedding, general $\longrightarrow$ $\mathtt{NEG}\ldots$
- Embed general metric into convex combination of tree metrics
  - Bartal's theorem: general $\longrightarrow$ distribution of a set of tree metrics.
  - Others: embed into one tree, embed into spanning trees.

Three major themes

- Embed general metric into norms
  - Bourgain's theorem: general $\rightarrow l_2$, distortion $O(\log n)$
  - Extension: volume respecting embedding, general $\rightarrow \mathtt{NEG} \ldots$
- Embed general metric into convex combination of tree metrics
  - Bartal's theorem: general $\rightarrow$ distribution of a set of tree metrics.
  - Others: embed into one tree, embed into spanning trees.
- Low-distortion embeddings between norms
  - JL-lemma $l_2^d \rightarrow l_2^k$ with $k = O(\log n/\epsilon^2)$ with $(1+\epsilon)$ distortion.

Applications

- Embed general metric into norms
  - Sparest cut/Multicommodity flow problem
  - Bandwidth problem

Applications

- Embed general metric into norms
  - Sparest cut/Multicommodity flow problem
  - Bandwidth problem
- Embed general metric into convex combination of tree metrics
  - Many online problems, metric task systems, distributed paging
  - Buy-at-bulk network design

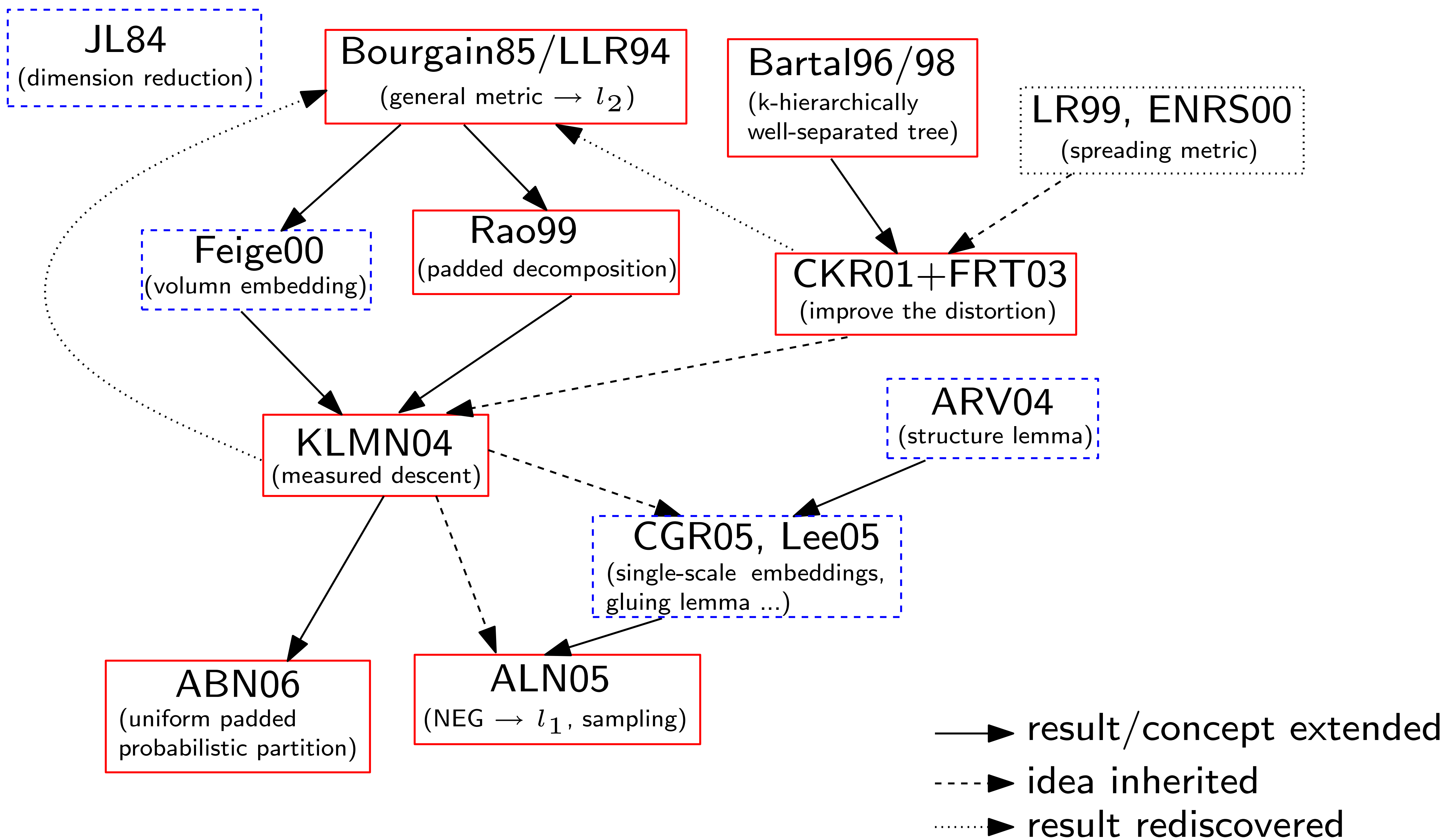
Applications

- Embed general metric into norms
  - Sparest cut/Multicommodity flow problem
  - Bandwidth problem
- Embed general metric into convex combination of tree metrics
  - Many online problems, metric task systems, distributed paging
  - Buy-at-bulk network design
- Low-distortion embeddings between norms
  - Approximating nearest neighbor
  - Clustering of high dimensional point set

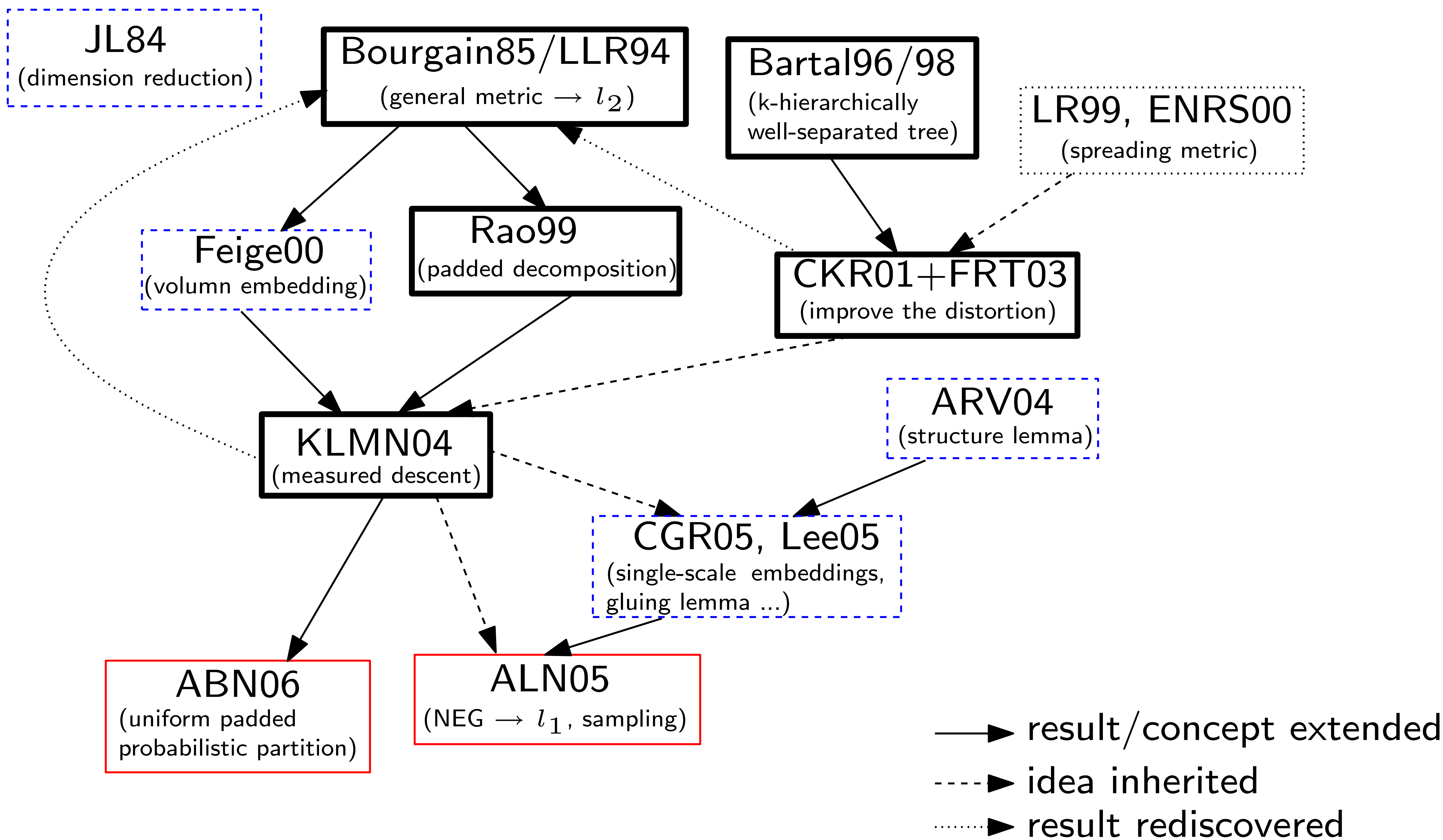
Three sources of power

- Power given by the metric
  - Triangle inequality
- Power given by random sampling and partition
  - Hierarchical structures: for all pairs of points in a particular input.
  - Randomization: for all the inputs
- Power given by random projection
  - Tight concentration property

Roadmap of the Techniques



Roadmap of the Techniques



Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

Bourgain's Embedding

- ## Bourgain's theorem

  Every finite n-point metric can be embedded into l_2 (And generally in l_p) with distortion $O(\log n)$.

Bourgain's Embedding

- Bourgain's theorem

  Every finite n-point metric can be embedded into l_2 (And generally in l_p) with distortion $O(\log n)$.
  Dimension of host space: $poly \log(n)$. (LLR94)

Bourgain's Embedding

- ## Bourgain's theorem

Every finite n-point metric can be embedded into l_2 (And generally in l_p) with distortion $O(\log n)$.
Dimension of host space: $poly \log(n)$. (LLR94)

- ## Fréchet-type embedding.

We define $f_Z(x) = d(x, Z) = \min_{s \in Z} d(x, s)$,
Z : zero set.

A *Fréchet-type embedding* is a map $f : X \rightarrow \mathbb{R}^k$ with

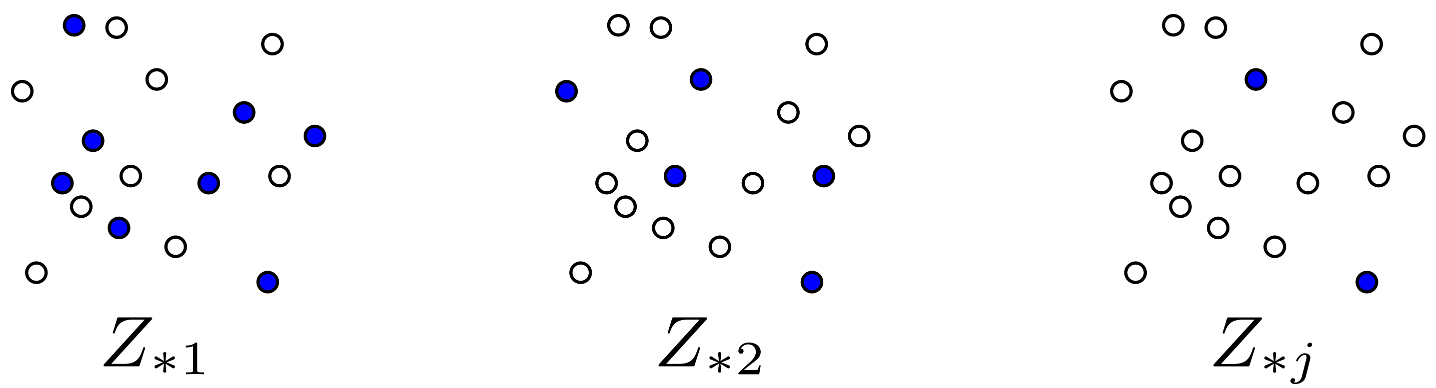
$$f(x) = \bigoplus_{Z \subseteq X} \beta_Z f_Z(x)$$

where $\beta_Z \in \mathbb{R}$.

Bourgain's Embedding Algorithm

Algorithm. Bourgain's embedding

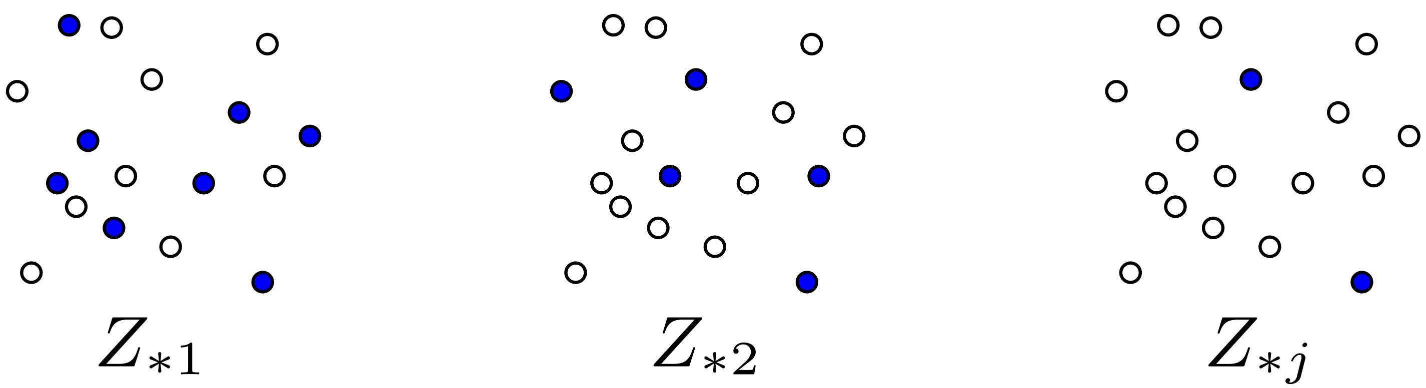
1. Pick random subsets Z_{ij} $(j = 1, 2, \ldots, \log n, i = 1, 2, \ldots, m = c \log n)$ with density levels 2^{-j}, (that is, each picked with Pr 2^{-j}).



Bourgain's Embedding Algorithm

- Algorithm. Bourgain's embedding

1. Pick random subsets Z_{ij} $(j = 1, 2, \ldots, \log n, i = 1, 2, \ldots, m = c \log n)$ with density levels 2^{-j}, (that is, each picked with Pr 2^{-j}).

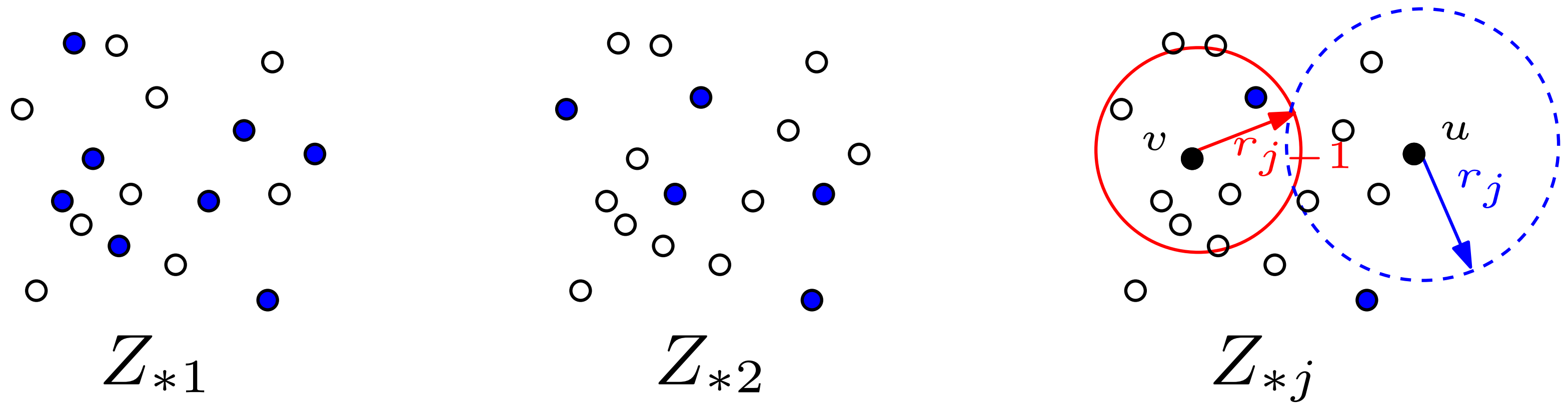


2. Fréchet-type embedding.
$f(x) = \bigoplus_{j=1}^{\log n} \bigoplus_{i=1}^{m} f_{ij}(x)$, with $f_{ij}(x) = d(x, Z_{ij})$.

Bourgain's Embedding Algorithm

- **Algorithm. Bourgain's embedding**

  1. Pick random subsets Z_{ij} $(j = 1, 2, \ldots, \log n, i = 1, 2, \ldots, m = c \log n)$ with density levels 2^{-j}, (that is, each picked with Pr 2^{-j}).



- The key observation:

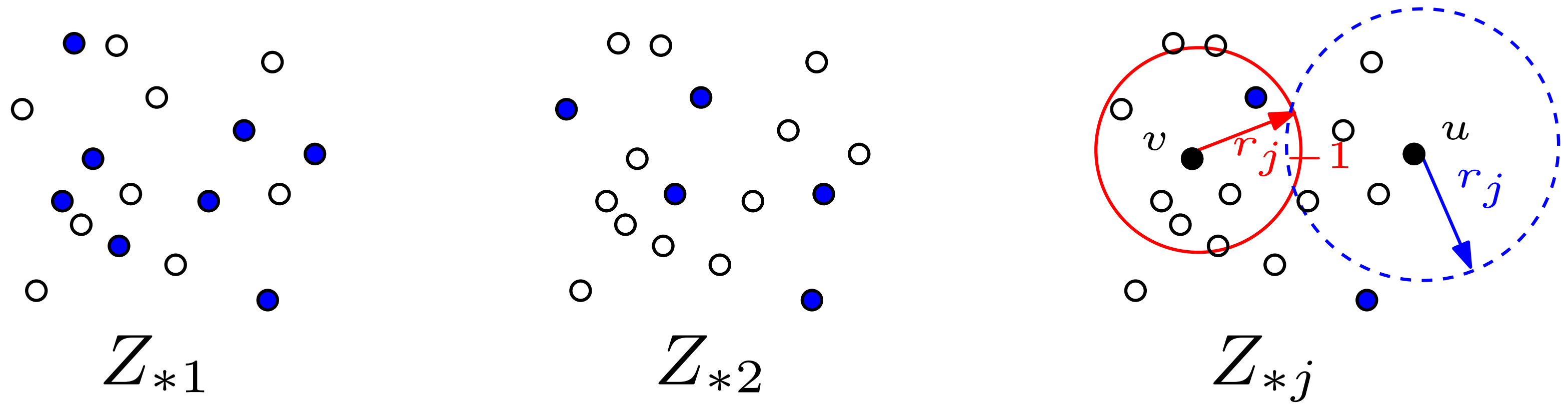
  $Pr((Z_{ij} \cap B(v, r_{j-1}) \neq \emptyset) \wedge (Z_{ij} \cap B^{\circ}(u, r_j) = \emptyset) \geq \Omega(1)$

  r_j : the smallest radius such that $|B(u, r_j)| \geq 2^j$ and $|B(v, r_j)| \geq 2^j$

Bourgain's Embedding Algorithm

- Algorithm. Bourgain's embedding

  1. Pick random subsets Z_{ij} $(j = 1, 2, \ldots, \log n, i = 1, 2, \ldots, m = c \log n)$ with density levels 2^{-j}, (that is, each picked with Pr 2^{-j}).



- The key observation:

  $Pr((Z_{ij} \cap B(v, r_{j-1}) \neq \emptyset) \wedge (Z_{ij} \cap B^{\circ}(u, r_j) = \emptyset) \geq \Omega(1)$

  r_j : the smallest radius such that $|B(u, r_j)| \geq 2^j$ and $|B(v, r_j)| \geq 2^j$

- We have $\|f_{ij}(u) - f_{ij}(v)\|_2 > r_j - r_{j-1}$.
  And then $\Omega(\log n) \cdot d(u, v) < f(u) - f(v) \leq O(\log^2 n) \cdot d(u, v)$.

Rao's padded decomposition - Algorithm

(planar metric $\rightarrow l_2$ with distortion $\sqrt{\log n}$)

- Algorithm. Rao's choice of zero sets (G, δ)

1. Pick an arbitrary node as root and build a BFS spanning tree.

2. $d(v)$ for a node v: its link-distance from the root in the BFS tree.
Choose a random number $r \in \{0, 1, \ldots, \delta\}$, and let S be the collection of all nodes with $d(v) \equiv r \pmod{\delta}$.

Rao's padded decomposition - Algorithm

(planar metric $\rightarrow l_2$ with distortion $\sqrt{\log n}$)

- Algorithm. Rao's choice of zero sets (G, δ)

1. Pick an arbitrary node as root and build a BFS spanning tree.

2. $d(v)$ for a node v: its link-distance from the root in the BFS tree.
Choose a random number $r \in \{0, 1, \ldots, \delta\}$, and let S be the collection of all nodes with $d(v) \equiv r \pmod{\delta}$.

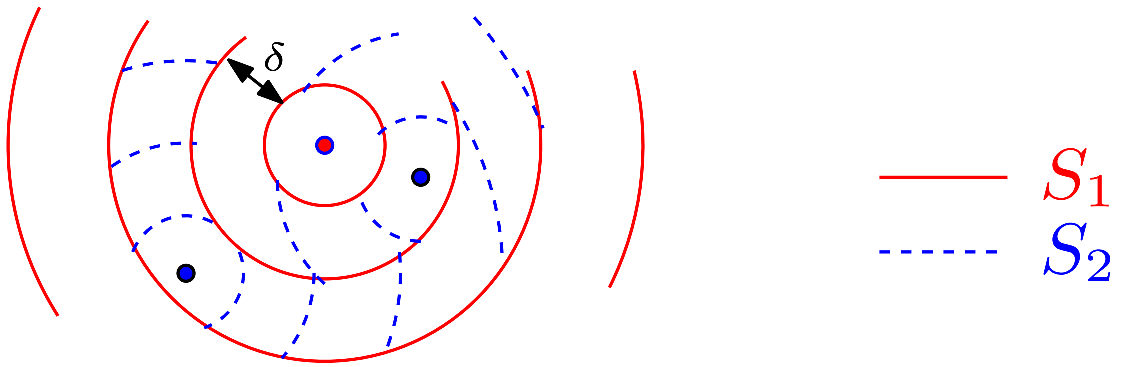
Rao's padded decomposition - Algorithm

(planar metric $\rightarrow l_2$ with distortion $\sqrt{\log n}$)

- Algorithm. Rao's choice of zero sets (G, δ)

1. Pick an arbitrary node as root and build a BFS spanning tree.

2. $d(v)$ for a node v: its link-distance from the root in the BFS tree.
Choose a random number $r \in \{0, 1, \ldots, \delta\}$, and let S be the collection of all nodes with $d(v) \equiv r \pmod{\delta}$.

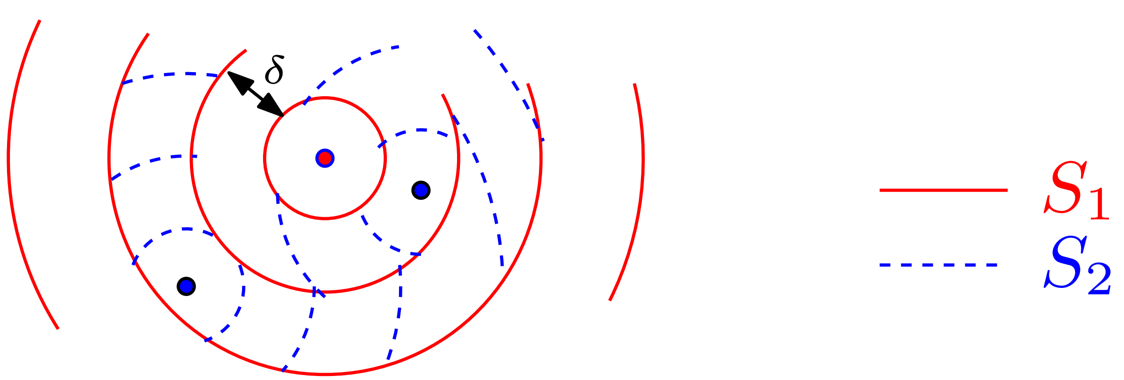


Rao's padded decomposition - Algorithm
(planar metric $\rightarrow l_2$ with distortion $\sqrt{\log n}$)

- Algorithm. Rao's choice of zero sets (G, δ)

1. Pick an arbitrary node as root and build a BFS spanning tree.

2. $d(v)$ for a node v: its link-distance from the root in the BFS tree.
Choose a random number $r \in \{0, 1, \ldots, \delta\}$, and let S be the collection of all nodes with $d(v) \equiv r \pmod{\delta}$.

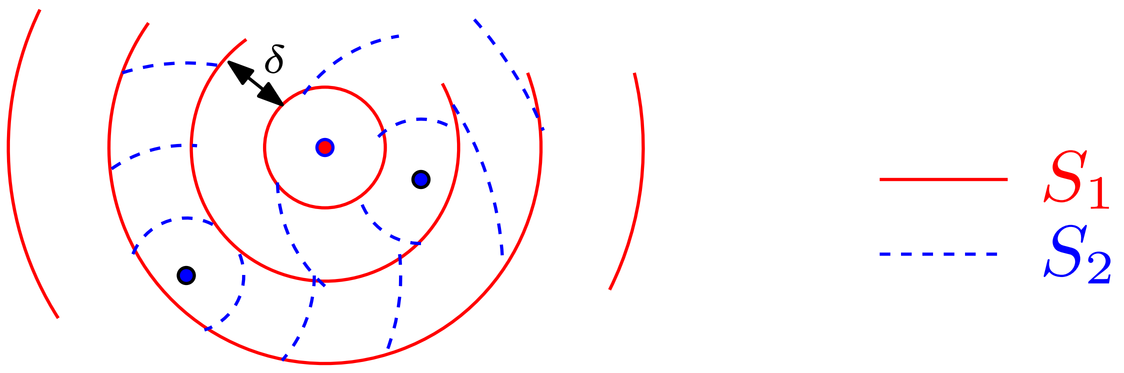


3. Decompose recursively 3 times.
Let $Z_\delta = S_1 \cup S_2 \cup S_3$

Rao's padded decomposition - Properties

- Properties of Rao's decomposition

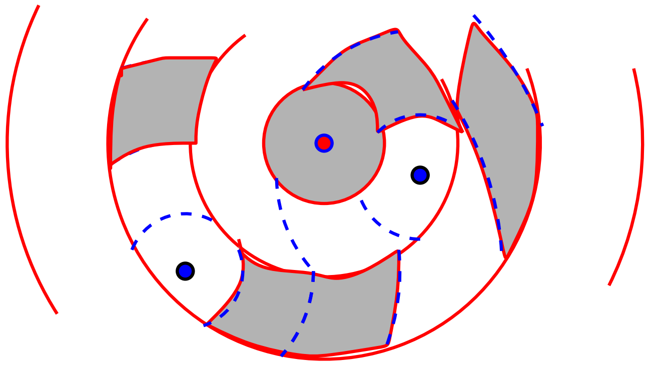
1. Each vertex is $\varepsilon\delta$ away from the boundary Z_δ w.c.p .
2. The diameter of each component of $G \setminus Z_\delta$ is at most $c\delta$.



Rao's padded decomposition - Embedding

- The embedding.

1. We assign a number $\sigma(C_i)$ chosen from symmetric (0,1) Bernoulli distribution for every component C_i.
2. For every $v \in C_i$, we create a coordinate $f_\delta(v) = \sigma(C_i)d(v, Z_\delta)$.



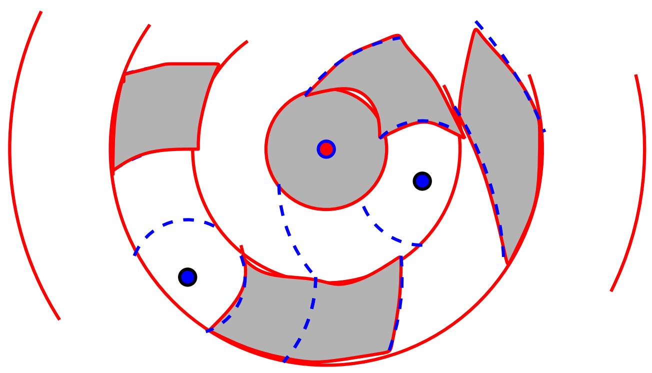
Rao's padded decomposition - Embedding

- The embedding.

  1. We assign a number $\sigma(C_i)$ chosen from symmetric (0,1) Bernoulli distribution for every component C_i.

  2. For every $v \in C_i$, we create a coordinate $f_\delta(v) = \sigma(C_i)d(v, Z_\delta)$.

  Key observation: w.c.p. $\|f_\delta(x) - f_\delta(y)\|_2 = \Omega(\delta)$.

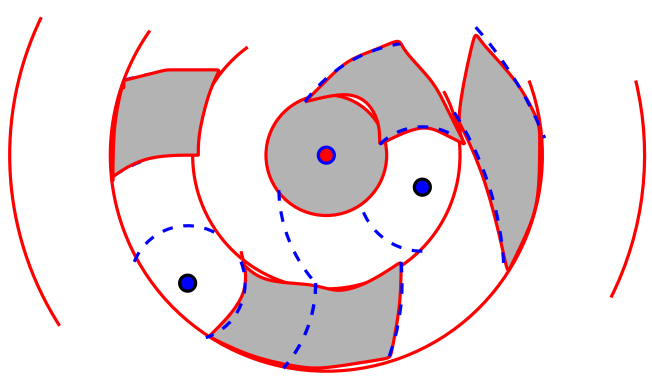


Rao's padded decomposition - Embedding

- The embedding.

1. We assign a number $\sigma(C_i)$ chosen from symmetric (0,1) Bernoulli distribution for every component C_i.
2. For every $v \in C_i$, we create a coordinate $f_\delta(v) = \sigma(C_i) d(v, Z_\delta)$.

   Key observation: w.c.p. $\|f_\delta(x) - f_\delta(y)\|_2 = \Omega(\delta)$.



3. Finally, $f(x) = \frac{1}{\beta \log n} \bigoplus_{\delta=2^t:1\leq 2^t \leq \Delta} \bigoplus_{i=1}^{\beta \log n} f_{i,\delta}(x)$.

Now we have, $\Omega(d(x,y)) < \|f(x) - f(y)\|_2 \leq O(\sqrt{\log n} \cdot d(x,y))$.

Rao's padded decomposition - Concepts

Metric decomposition

- (X, d): a finite metric space.
- A decomposition $P = \{C_1, C_2, \ldots, C_m\}$: a partition decomposing a graph to several components by deleting some edges (or vertices).
- $P(x)$: the component containing x.
- $\mathcal{P}_X$: the collection of all decompositions of X.
- $\mathcal{D}_X$: the set of all probability distributions on $\mathcal{P}_X$.

Rao's padded decomposition - Concepts

Padded decomposition

A random decomposition of a finite metric space (X, d) is a distribution $Pr \in \mathcal{D}_X$ over decompositions of X.

For a particular scale $\delta > 0$ and a function $\varepsilon : X \to (0, 1]$, a δ-bounded ε-padded decomposition is one which satisfies the following two properties.

1. For all $P \in \text{supp}(Pr)$, $\texttt{diam}(C) \leq \delta$ for all $C \in P$.
2. For all $x \in X$, $Pr[B(x, \varepsilon(x)\delta) \subseteq P(x)] \geq 1/2$.

Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

Bartal's theorem

- The theorem

  (X, d): a finite n-point metric let $\mathcal{T}$ be the set of all tree metrics that dominate d.

  There is a distribution π on $\mathcal{T}$, $r > 0$ and $D = O(\log^2 n)$ such that pick $T \in \mathcal{T}$ randomly according to π

$$\forall x, y \in X, \quad r \leq \frac{E[d_T(x,y)]}{d_X(x,y)} \leq Dr$$

Bartal's theorem

- The theorem

  (X, d): a finite n-point metric let $\mathcal{T}$ be the set of all tree metrics that dominate d.

  There is a distribution π on $\mathcal{T}$, $r > 0$ and $D = O(\log^2 n)$ such that pick $T \in \mathcal{T}$ randomly according to π

$$\forall x, y \in X, \quad r \leq \frac{E[d_T(x,y)]}{d_X(x,y)} \leq Dr$$

- Approach.

  Create k-hierarchically well-separated trees (k-HST).

Bartal's theorem - Construction of k-HST

Algorithm Construction of k-HST (G, i)

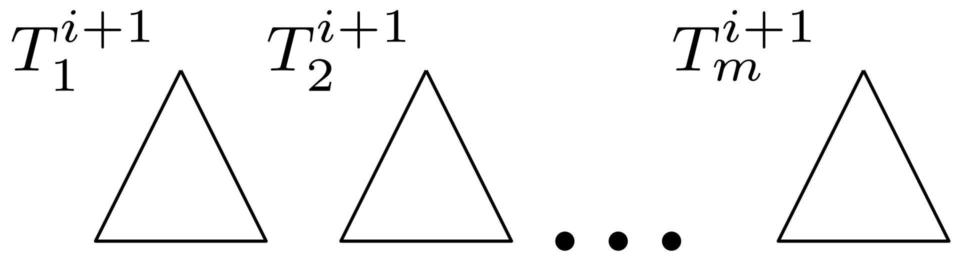
1. If G is a single vertex, return G.

2. $E' =$ **Bartal's probabilistic decomposition**$(G, \Delta/k^{i+1})$
Let $G_1^{i+1}, G_2^{i+1}, \ldots, G_m^{i+1}$ be the components left by deleting E' from G.

Bartal's theorem - Construction of k-HST

Algorithm Construction of k-HST (G, i)

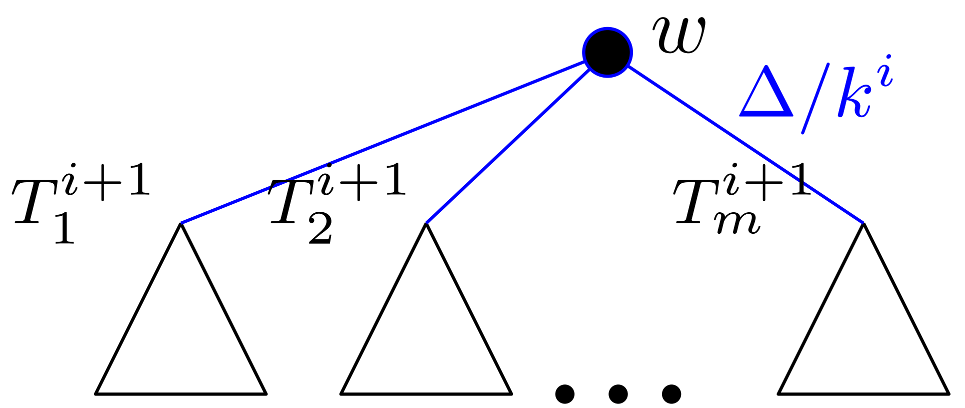
1. If G is a single vertex, return G.
2. $E' =$ **Bartal's probabilistic decomposition**$(G, \Delta/k^{i+1})$
Let $G_1^{i+1}, G_2^{i+1}, \ldots, G_m^{i+1}$ be the components left by deleting E' from G.
3. For $j = 1$ to m, $T_j^{i+1} =$ construct_k-HST$(G_j^{i+1}, i+1)$.



Bartal's theorem - Construction of k-HST

Algorithm Construction of k-HST (G, i)

1. If G is a single vertex, return G.
2. E' = **Bartal's probabilistic decomposition**$(G, \Delta/k^{i+1})$
Let $G_1^{i+1}, G_2^{i+1}, \ldots, G_m^{i+1}$ be the components left by deleting E' from G.
3. For $j = 1$ to m, $T_j^{i+1} =$ construct_k-HST$(G_j^{i+1}, i+1)$.
4. Create a new root w and glue the $T_j^{i+1} (j = 1, 2, \ldots, m)$ as subtrees of w by using edges with length Δ/k^i, thus form a new tree T, return T.



Bartal's theorem - Construction of k-HST

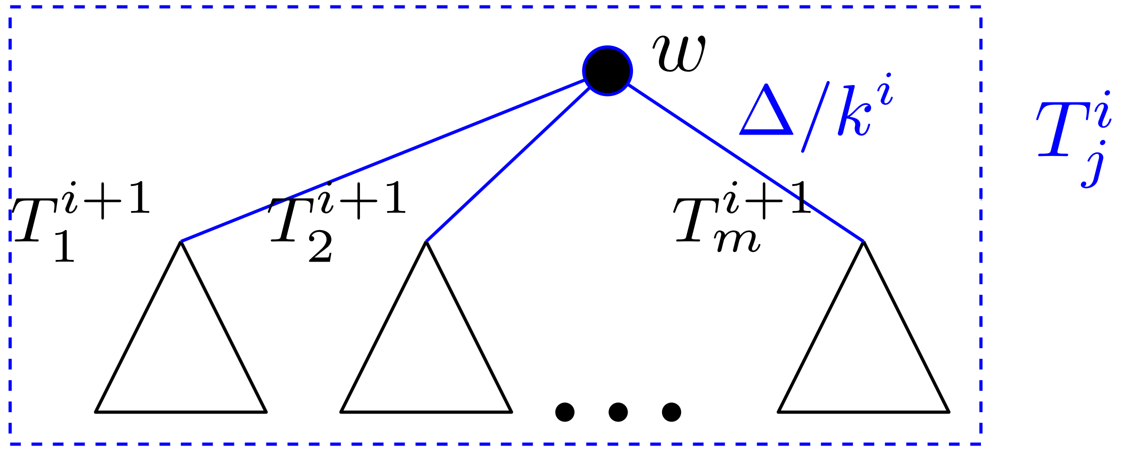
Algorithm Construction of k-HST (G, i)

1. If G is a single vertex, return G.

2. $E' =$ **Bartal's probabilistic decomposition**$(G, \Delta/k^{i+1})$
Let $G_1^{i+1}, G_2^{i+1}, \ldots, G_m^{i+1}$ be the components left by deleting E' from G.

3. For $j = 1$ to m, $T_j^{i+1} =$ construct_k-HST$(G_j^{i+1}, i+1)$.

4. Create a new root w and glue the $T_j^{i+1} (j = 1, 2, \ldots, m)$ as subtrees of w by using edges with length Δ/k^i, thus form a new tree T, return T.



Bartal's theorem - Properties

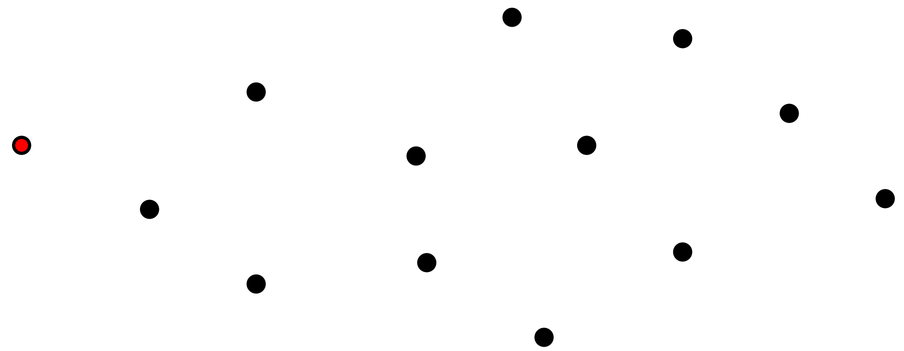
- Properties of the decomposition (2-HST)

1. Every connected component in $G \setminus E'$ has diameter no more than δ.
2. For every $e \in E$, $Pr[\text{edge } e \text{ is deleted}] \leq 4 \log n \cdot (d(e)/\delta)$.

Bartal's theorem - Algorithm

Algorithm. Bartal's probabilistic decomposition (G, δ)

- Let $G' = G$, and $E' = \emptyset$
- While $G' \neq \emptyset$ do

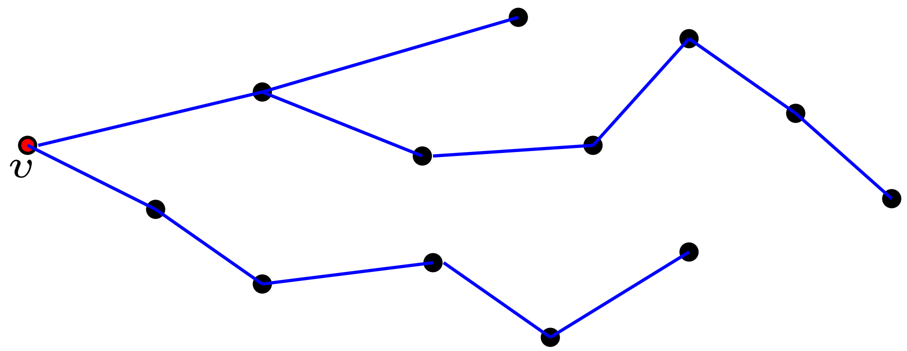


Bartal's theorem - Algorithm

Algorithm. Bartal's probabilistic decomposition (G, δ)

- Let $G' = G$, and $E' = \emptyset$
- While $G' \neq \emptyset$ do

  1. Pick an arbitrary vertex $v \in G'$ and grow a shortest path tree in G' rooted at v. Let $F = \emptyset$.

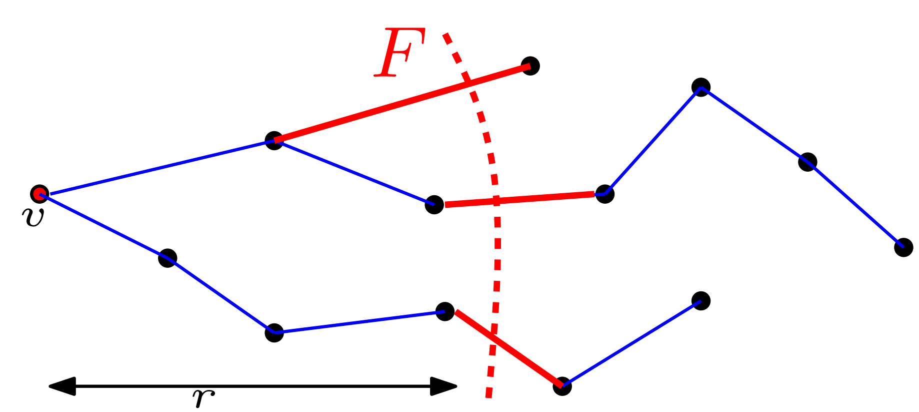


Bartal's theorem - Algorithm

Algorithm. Bartal's probabilistic decomposition (G, δ)

- Let $G' = G$, and $E' = \emptyset$
- While $G' \neq \emptyset$ do

1. Pick an arbitrary vertex $v \in G'$ and grow a shortest path tree in G' rooted at v. Let $F = \emptyset$.
2. Let r be a random variable drawn from geometric distribution with parameter $p = \frac{4 \log n}{\delta}$.
3. Add to F all edges that cross $[r-1, r]$ in the shortest path tree.



Bartal's theorem - Algorithm

Algorithm. Bartal's probabilistic decomposition (G, δ)

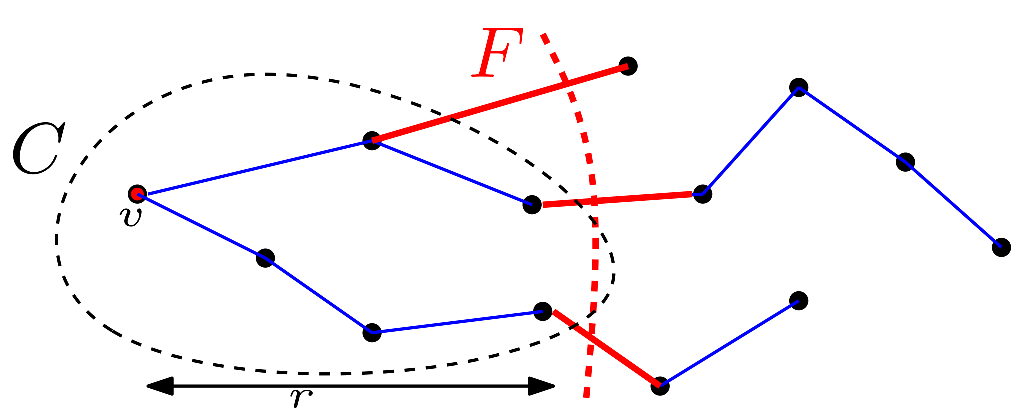
- Let $G' = G$, and $E' = \emptyset$
- While $G' \neq \emptyset$ do

  1. Pick an arbitrary vertex $v \in G'$ and grow a shortest path tree in G' rooted at v. Let $F = \emptyset$.

  2. Let r be a random variable drawn from geometric distribution with parameter $p = \frac{4 \log n}{\delta}$.

  3. Add to F all edges that cross $[r-1, r]$ in the shortest path tree.

  4. Let C denote the component containing v after F is removed. Let $G' = G' \setminus C$. Let $E' = E' \cup F$.



Bartal's theorem - Proof of the properties

Proof of the two properties:

- Probability of cut is proportional to the length

$$\begin{aligned} Pr[\text{edge } e \text{ is deleted}] &\leq 1-(1-p)^{d(e)} \\ &\leq 1-(1-pd(e)) \\ &= \frac{4\log n}{\delta} d(e) \end{aligned}$$

Bartal's theorem - Proof of the properties

Proof of the two properties:

- Probability of cut is proportional to the length

$$
\begin{aligned}
Pr[\text{edge } e \text{ is deleted}] &\leq 1-(1-p)^{d(e)} \\
&\leq 1-(1-pd(e)) \\
&= \frac{4\log n}{\delta} d(e)
\end{aligned}
$$

- Small Diameter

$$
\begin{aligned}
Pr[\mathtt{diam}(C) > \delta] &\leq Pr[r > \delta/2] \\
&= (1-p)^{\delta/2} \\
&\leq e^{-p\frac{\delta}{2}} = \frac{1}{n^2}
\end{aligned}
$$

Bartal's theorem - Analysis

Analysis. (Invoke the above algo with $\delta = \Delta/2^{j+1}$)

Fact from the second property:

$$Pr[x, y \text{ lie in two } T^j\text{s} \mid x, y \text{ lie in a same } T^{j-1}] \leq 4 \log n \cdot \frac{d(x,y)}{\Delta/2^{j+1}}$$

Bartal's theorem - Analysis

Analysis. (Invoke the above algo with $\delta = \Delta/2^{j+1}$)

Fact from the second property:

$$Pr[x, y \text{ lie in two } T^j\text{s} \mid x, y \text{ lie in a same } T^{j-1}] \leq 4 \log n \cdot \frac{d(x,y)}{\Delta/2^{j+1}}$$

And then

$$\begin{aligned} E_R[d_T(x,y)] &= 4\Delta \cdot Pr[x \text{ and } y \text{ lie in two different } T^1\text{s}] \\ &+ \sum_{j=1}^{\log(4\Delta)} \frac{4\Delta}{2^j} \cdot Pr[\text{they lie in two } T^{j+1}\text{s} \mid \\ &\quad \text{they lie in a same } T^j] \\ &\leq 32 \log n \log \Delta \cdot d(x,y) \end{aligned}$$

Bartal's theorem - Analysis

Analysis. (Invoke the above algo with $\delta = \Delta/2^{j+1}$)

Fact from the second property:

$$Pr[x, y \text{ lie in two } T^j\text{s} \mid x, y \text{ lie in a same } T^{j-1}] \leq 4 \log n \cdot \frac{d(x,y)}{\Delta/2^{j+1}}$$

And then

$$\begin{aligned} E_R[d_T(x,y)] &= 4\Delta \cdot Pr[x \text{ and } y \text{ lie in two different } T^1\text{s}] \\ &+ \sum_{j=1}^{\log(4\Delta)} \frac{4\Delta}{2^j} \cdot Pr[\text{they lie in two } T^{j+1}\text{s} \mid \\ &\qquad \text{they lie in a same } T^j] \\ &\leq 32 \log n \log \Delta \cdot d(x,y) \qquad \Rightarrow O(\log^2 n) \end{aligned}$$

Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

FRT's analysis

- How to save a $\log n$ factor?
  - We try to perform a random cut in a more restrictive range.
  - Concretely, we perform a multi-scale decomposition: we choose a sequence of widths of ranges (i.e. $[\delta, 2\delta], [2\delta, 4\delta] \ldots$) instead of using a sequence of geometrically distributed random variables in a same range to perform the decomposition.

FRT's analysis

- How to save a $\log n$ factor?
  - We try to perform a random cut in a more restrictive range.
  - Concretely, we perform a multi-scale decomposition: we choose a sequence of widths of ranges (i.e. $[\delta, 2\delta], [2\delta, 4\delta] \ldots$) instead of using a sequence of geometrically distributed random variables in a same range to perform the decomposition.
- Use decomposition method to prove Bourgain's theorem (l_1 case).

FRT's decomposition - Algorithm

- Algorithm. FRT's decomposition (X, δ)

1. Choose uniformly at random $R \in [\delta/2, \delta]$, and an ordering π on the elements of X.

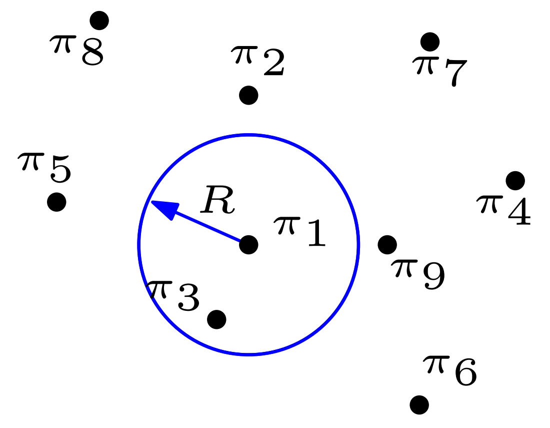
2. For each element in $x \in X$ according to the ordering π, we collect all the elements within distance R from x that are not marked before, form a new component and mark them.

FRT's decomposition - Algorithm

- Algorithm. FRT's decomposition (X, δ)

  1. Choose uniformly at random $R \in [\delta/2, \delta]$, and an ordering π on the elements of X.

  2. For each element in $x \in X$ according to the ordering π, we collect all the elements within distance R from x that are not marked before, form a new component and mark them.

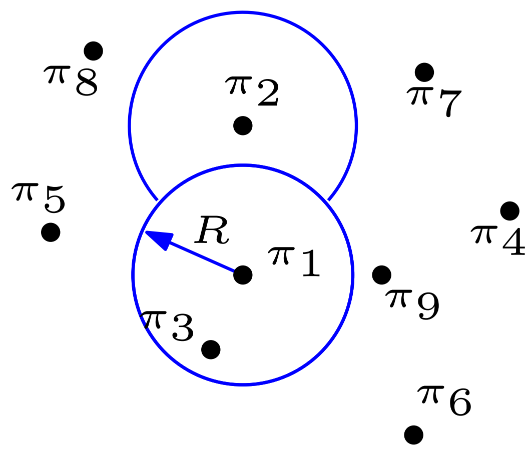


FRT's decomposition - Algorithm

- **Algorithm. FRT's decomposition (X, δ)**

  1. Choose uniformly at random $R \in [\delta/2, \delta]$, and an ordering π on the elements of X.

  2. For each element in $x \in X$ according to the ordering π, we collect all the elements within distance R from x that are not marked before, form a new component and mark them.

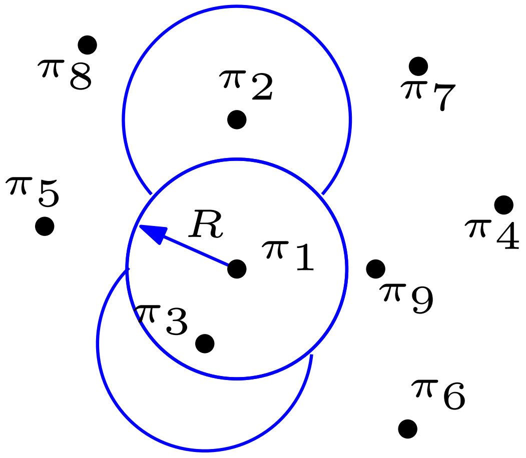


FRT's decomposition - Algorithm

- **Algorithm. FRT's decomposition (X, δ)**

  1. Choose uniformly at random $R \in [\delta/2, \delta]$, and an ordering π on the elements of X.

  2. For each element in $x \in X$ according to the ordering π, we collect all the elements within distance R from x that are not marked before, form a new component and mark them.

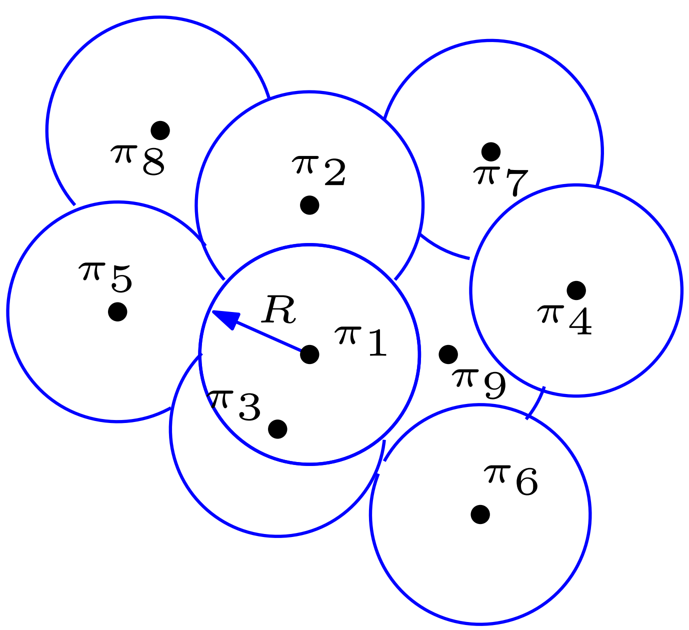


FRT's decomposition - Algorithm

- **Algorithm. FRT's decomposition (X, δ)**

1. Choose uniformly at random $R \in [\delta/2, \delta]$, and an ordering π on the elements of X.

2. For each element in $x \in X$ according to the ordering π, we collect all the elements within distance R from x that are not marked before, form a new component and mark them.



FRT's analysis - Properties

- Properties of the decomposition.

1. Every component left has diameter between $\delta/2$ and δ.
2. For every $x \in X$,

$$Pr[\texttt{Ball}(x,\tau) \not\subseteq P(x)] \leq \frac{4\tau}{\delta/2} \log \frac{|\texttt{Ball}(x,\delta+\tau)|}{|\texttt{Ball}(x,\delta/2-\tau)|}$$

FRT's analysis - Properties

- Properties of the decomposition.

  1. Every component left has diameter between $\delta/2$ and δ.
  2. For every $x \in X$,

$$Pr[\texttt{Ball}(x,\tau) \not\subseteq P(x)] \le \frac{4\tau}{\delta/2} \log \frac{|\texttt{Ball}(x,\delta+\tau)|}{|\texttt{Ball}(x,\delta/2-\tau)|}$$

- Consequently.

$$Pr\left[\texttt{Ball}(x, \frac{\delta}{16\log n}) \subseteq P(x)\right] \ge 1/2$$

since $\frac{|\texttt{Ball}(x,2\delta)|}{|\texttt{Ball}(x,\delta/4)|} \le n$.

FRT's analysis - Analysis

Analysis

E_v be the events that v is the first element chosen in the ordering π such that $d(x, v) < R + \tau$.

$\texttt{Ball}(x, \tau) \not\subseteq P(x) \Rightarrow \bigvee_{v:d(x,v)\in[R-\tau,R+\tau]} E_v$.

To bound the probability of the left hand side, it suffices to bound the probability of the E_v.

FRT's analysis - Analysis

Analysis

E_v be the events that v is the first element chosen in the ordering π such that $d(x,v) < R + \tau$.

$$\texttt{Ball}(x,\tau) \not\subseteq P(x) \Rightarrow \bigvee_{v:d(x,v)\in[R-\tau,R+\tau]} E_v.$$

To bound the probability of the left hand side, it suffices to bound the probability of the E_v.

Let $v_1, v_2, \ldots, v_i$ be the points of the set.

$$\begin{aligned} Pr[E_{v_i}] &\leq Pr[R \in [d(x,v_i)-\tau, d(x,v_i)+\tau] \wedge \\ &\qquad v_i \text{ is the 1st element chosen with } d(x,v_i) < R+\tau \\ &\leq \frac{2\tau}{\delta/2} \cdot \frac{1}{|\texttt{Ball}(x, R+\tau)|} \end{aligned}$$

FRT's analysis - Analysis

Analysis

E_v be the events that v is the first element chosen in the ordering π such that $d(x,v) < R+\tau$.

$$\texttt{Ball}(x,\tau) \not\subseteq P(x) \Rightarrow \bigvee_{v:d(x,v)\in[R-\tau,R+\tau]} E_v.$$

To bound the probability of the left hand side, it suffices to bound the probability of the E_v.

Let $v_1, v_2, \ldots, v_i$ be the points of the set.

$$\begin{aligned} Pr[E_{v_i}] &\leq Pr[R \in [d(x,v_i)-\tau, d(x,v_i)+\tau] \wedge \\ &\quad v_i \text{ is the 1st element chosen with } d(x,v_i) < R+\tau \\ &\leq \frac{2\tau}{\delta/2} \cdot \frac{1}{|\texttt{Ball}(x,R+\tau)|} \end{aligned}$$

Then integration

FRT's analysis - Embedding

- Embedding procedure.
  - Z_δ: the union of boundaries of all components.

  1. $f_\delta(x) = \sigma(C_i) d(x, Z_\delta)$. $\sigma(C_i)$ as before.
  2. $f(x) = \bigoplus_\delta f_\delta(x)$.

- And then, recall Rao's embedding ...

Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

The measured descent technique

- The intuition.

  Given a fix point x, and for another point y, we hope that there is a random decomposition with scale $\lfloor d(x,y) \rfloor$, which will provide considerable probability that x, y is well separated in that coordinate.

The measured descent technique

- The intuition.

  Given a fix point x, and for another point y, we hope that there is a random decomposition with scale $\lfloor d(x,y) \rfloor$, which will provide considerable probability that x, y is well separated in that coordinate.

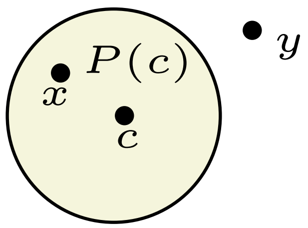
- A naive solution

  For each point v, we choose a (near) optimal scale series of cardinality $O(\log n)$ according to the distance between v and other points, instead of choosing the same geometrically decreasing scale series for all the points.

The measured descent technique -Difficulty

- The difficulty.

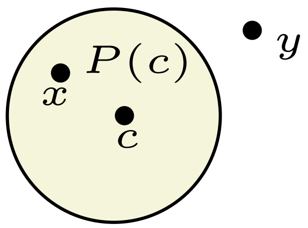
  If we still use the previous decomposition method, the scale series we chosen for c might not be good for x, the local consistency is lost.



The measured descent technique -Difficulty

- The difficulty.

  If we still use the previous decomposition method, the scale series we chosen for c might not be good for x, the local consistency is lost.



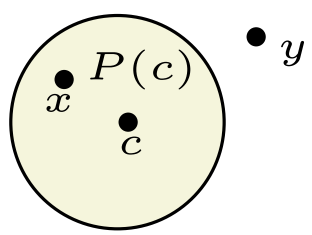
- Solution

  Create subscales to smooth the scale series.

The measured descent technique -Difficulty

- The difficulty.

  If we still use the previous decomposition method, the scale series we chosen for c might not be good for x, the local consistency is lost.



- Solution

  Create subscales to smooth the scale series.

- Underlying observation.

  Low density variation of the neighborhood.

Outline

- Introduction
- Themes and sources of power
- Review
- Bartal's recursive probabilistic decomposition
- FRT's analysis
- The measure descent technique
- Conclusion

Summary

Where is the power?

- The hierarchical structure makes it possible that a certain embedding is good for all pairs of points in a particular input.

- Randomization gives us the power to prove that a certain hierarchical structure would be good for all inputs with, say, some constant probability.

Summary

Where is the power?

- The hierarchical structure makes it possible that a certain embedding is good for all pairs of points in a particular input.
- Randomization gives us the power to prove that a certain hierarchical structure would be good for all inputs with, say, some constant probability.
- The combination of the two provides many of the results in the metric embedding paradigm.

The End

THANK YOU

Q and A