Tree Metrics[1]

R. Inkulu
http://www.iitg.ac.in/rinkulu/

[1] slides last updated in 2013

Outline

Metric Space: Definition

A *metric space* is a pair (X, d) where X is a set and $d : X \times X \to [0, \infty)$ is a metric satisfying:

- $d_{xy} \geq 0$
- $d_{xy} = 0$ iff $x = y$
- $d_{xy} = d_{yx}$
- $d_{xy} + d_{yz} \geq d_{xz}$ (triangle inequality)

ex. $\mathcal{R}^d$

Finite Metric Space: Definition

A metric space (X, d) is a *finite metric space* if $|X|$ is finite.

ex. graph metric a.k.a. metric completion of a graph

Any finite metric space can be represented by a complete weighted graph

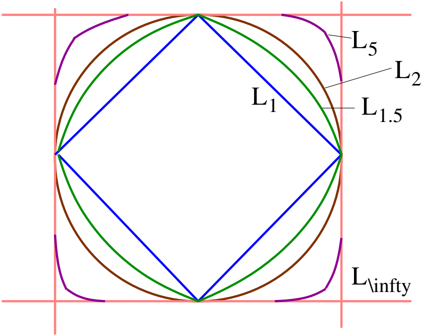
L_p^d-Minkowski norms

For $p \geq 1$, L_p^d defines the distance between two points $x, y \in R^d$ as $\|x - y\|_p^d = (\sum_{i=1}^d |x_i - y_i|^p)^{1/p}$. The popular norms include:

- rectilinear norm L_1^d: $\|x - y\|_1 = \sum_{i=1}^d |x_i - y_i|$
- Euclidean norm L_2^d: $\|x - y\|_2 = \sum_{i=1}^d |x_i - y_i|^2$
- max norm L_∞^d: $\|x - y\|_\infty = \max_{i=1}^d |x_i - y_i|$

The triangle inequality holds for all Minkowski norms.

Unit balls in various Minkowski norms



- For any $p \in \mathcal{R}^d$, $\frac{\|p\|_1}{\sqrt{d}} \leq \|p\|_2 \leq \|p\|_1$

Embedding

Let (X, d') and (Y, d'') be two (finite) metric spaces. Any one-to-one map $f : X \rightarrow Y$ is termed an *embedding*.

An embedding that preserves the distance between every two points is termed an *isometric embedding*.

An embedding in which no distance shrinks is termed an *expansive embedding*.

Distortion of an embedding

- The mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is called *K-bi-Lipschitz* for a subset $X \subseteq \mathbb{R}^n$ if there exists a constant $c > 0$ such that
  $cK^{-1}\|p - q\| \leq \|f(p) - f(q)\| \leq c\|p - q\|$.
  The least K for which f is K-bi-Lipschitz is called the *distortion* of f.

Distortion of an embedding

- The mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is called *K-bi-Lipschitz* for a subset $X \subseteq \mathbb{R}^n$ if there exists a constant $c > 0$ such that
  $cK^{-1}\|p - q\| \leq \|f(p) - f(q)\| \leq c\|p - q\|$.
  The least K for which f is K-bi-Lipschitz is called the *distortion* of f.

- Let (X, d') and (Y, d'') be two (finite) metric spaces. The distortion of an expansive embedding $f : X \to Y$ is $\max_{x,y \in X} \frac{d''(f(x), f(y))}{d'(x,y)}$.

Algorithm design: typical pipeline

- viewing a combinatorial problem (ex. shortest paths in graphs) as a finite metric space, say (V, d)
- embed (V, d) into a finite metric space (V', d')
- using an efficient algorithm, solve the problem in (V', d')

distortion corresp. to the function that embeds is an apprx factor

Outline

Advantages of tree metrics

- for many problems, efficient algorithms are available for trees
- trees are embeddable into L_1 with no distortion

Cycle to tree embedding

Embedding a unit weighted cycle C into a tree T while T being a subgraph of C:

- lower bound stands at $\Omega(n)$

Tree Metric: definition

An α-distortion embedding of a finite metric space (V, d) into a *tree* (V', T)[2]:

- $V \subseteq V'$
- positive edge length associated to each edge of T
- $d_{uv} \leq T_{uv} \leq \alpha d_{uv}$

further, if $V = V'$, then (V', T) is termed as a *spanning tree metric*.

w.l.o.g., $d_{uv} \geq 1$ for all $u \neq v$ in V

[2] with a few minor adjustments (moving labels on edges to nodes), the tree to be constructed is a *hierarchically well-separated tree*

Embedding cycle into tree metric

Embedding a unit weighted cycle C into a tree T while T being a supergraph of C:

- again, the lower bound stands at $\Omega(n)$ — not proved

how about *probabilistically embedding*?

Expected distortion

Probabilistically embedding a metric space (X, d) into convex combination of trees:

- Let $T_1, T_2, \ldots, T_k$ be a sequence of metrics $T_i : (X, d_i)$, and let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be positive reals with $\sum_i \alpha_i = 1$. Then Ts and αs together define a *probabilistic metric*. The expected distance between p and q is $\sum_i \alpha_i d_i(p, q)$.

The expected distortion is $max_{u,v \in V} \frac{E[T(u,v)]}{d'_{uv}}$.

Significant results

Bartal's result ([Bartal96]) devised a randomized polynomial time algorithm for the following:

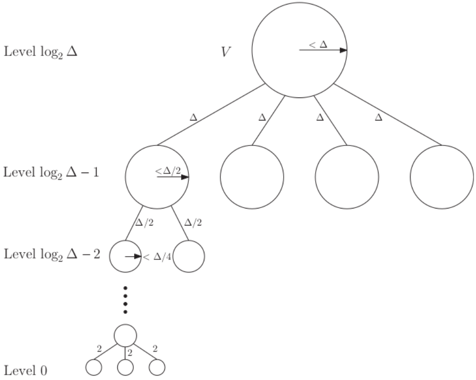
for $(|X| = n, d')$ being an arbitrary metric, and X' is finite,

$(X, d') \hookrightarrow^{O(\lg^2 n) expected} (X', T)$ for $X \subseteq X'$

Later the expected distortion got improved to $O(\lg n)$ due to a randomized polynomial time algorithm devised in [FRT04].

The lower bound on the expected distortion in probabilistically approximating metrics by tree metrics is known to be $\Omega(\lg n)$. — not proved

Outline



for a given node at level i that corresp. to a set S, vertices in S will be the vertices in a ball of radius $< 2^i$ and $\geq 2^{i-1}$ centered at some vertex [3]

here, $\Delta = \min_x 2^x > 2 \cdot max_{u,v \in V} d_{uv}$

- root has the entire V; each leaf node corresp. to a unique point in V
- nodes in each level together partition V

[3] vertices of T are referred as nodes while the vertices of V are referred as points

Hierarchical cut decomposition is a tree metric

(V', T) is an expansive tree metric embedding of (V, d) due to:

- $V \subseteq V'$
- positive edge lengths
- (V', T) is an expansive metric

  lowest level at which u and v belong to the same is $\lfloor \lg_2 d_{uv} \rfloor$
- what about the distortion?

Randomized Algorithm to construct (V', T)

1. pick a permutation π of V
2. pick a random number r_0 in $[1/2, 1)$; set radius $r_i = 2^i r_0$ for all balls at each level i
3. root is associated with points in ball $B(\text{any point}, \Delta)$ i.e., V itself
4. for each node v in each level i ($i > 0$)

   let S be the set of points associated with v

   (a) for each j from 1 to n

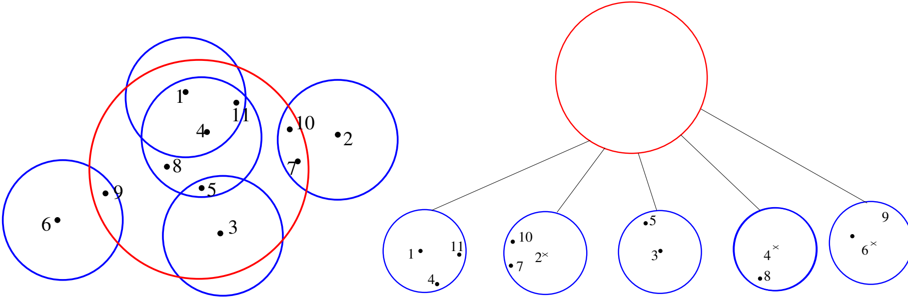
      (i) if $S' = B(\pi(j), r_{i-1}) \cap S \neq \phi$ then create a child node to v and associate points in S' to it

      (ii) $S = S - S'$

   (b) for each edge e that got created in (a), set the weight of e to 2^i

takes randomized polynomial time

Algorithm in Execution



points belonging to a tree node are shown with filled circles

Outline

lower bound on T_{uv}

$\forall_{u,v \in V}\, d_{uv} \leq T_{uv}$

- if u and v are in a set S corresp. to a node at level i, then $d_{uv} < 2^{i+1}$ (since the radius of the ball containing S is $< 2^i$)

  hence, u and v cannot belong to the same node at level $\lfloor \lg_2 d_{uv} \rfloor - 1$

  implies, the lowest level at which u and v can belong to the same node is $\lfloor \lg_2 d_{uv} \rfloor$
- therefore, the distance $T_{uv} \geq 2 \sum_{j=1}^{\lfloor \lg_2 d_{uv} \rfloor} 2^j \geq d_{uv}$

Upper bounding the expected distortion

- If LCA of u and v is at level i, then $T_{uv} \leq 2^{i+2}$.
  - $T_{uv} = 2\sum_{j=1}^{i} 2^j = 2^{i+2} - 4 \leq 2^{i+2}$

Upper bounding the expected distortion

- If LCA of u and v is at level i, then $T_{uv} \leq 2^{i+2}$.
  - $T_{uv} = 2\sum_{j=1}^{i} 2^j = 2^{i+2} - 4 \leq 2^{i+2}$

- $E[T_{uv}]$

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} \text{pr}($
  w is the first vertex in the random permutation of vertices such that at least one of u, v is in the ball $B(w, r_i) \quad \wedge$
  exactly one of u and v is in $B(w, r_i)$
  $) * (T_{uv}$ when the LCA of u and v is in level $i + 1)$

Upper bounding the expected distortion

- If LCA of u and v is at level i, then $T_{uv} \leq 2^{i+2}$.
  - $T_{uv} = 2\sum_{j=1}^{i} 2^j = 2^{i+2} - 4 \leq 2^{i+2}$

- $E[T_{uv}]$

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1}$ pr(
  w is the first vertex in the random permutation of vertices such that at least one of u, v is in the ball $B(w, r_i)$ $\wedge$
  exactly one of u and v is in $B(w, r_i)$
  $) * (T_{uv}$ when the LCA of u and v is in level $i+1)$

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} pr(S_{iw} \wedge X_{iw}) 2^{i+3}$ (respectively denoted the above two descriptions with S_{iw} and X_{iw})

Upper bounding the expected distortion

- If LCA of u and v is at level i, then $T_{uv} \leq 2^{i+2}$.
  - $T_{uv} = 2\sum_{j=1}^{i} 2^j = 2^{i+2} - 4 \leq 2^{i+2}$

- $E[T_{uv}]$

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1}$ pr(
  w is the first vertex in the random permutation of vertices such that at least one of u, v is in the ball $B(w, r_i)$ $\wedge$
  exactly one of u and v is in $B(w, r_i)$
) $*$ (T_{uv} when the LCA of u and v is in level $i+1$)

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} pr(S_{iw} \wedge X_{iw}) 2^{i+3}$ (respectively denoted the above two descriptions with S_{iw} and X_{iw})

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} pr(S_{iw} | X_{iw}) pr(X_{iw}) 2^{i+3}$

Upper bounding the expected distortion

- If LCA of u and v is at level i, then $T_{uv} \leq 2^{i+2}$.
  - $T_{uv} = 2\sum_{j=1}^{i} 2^j = 2^{i+2} - 4 \leq 2^{i+2}$
- $E[T_{uv}]$

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1}$ pr(
  w is the first vertex in the random permutation of vertices such that at least one of u, v is in the ball $B(w, r_i)$ $\wedge$
  exactly one of u and v is in $B(w, r_i)$
) $* \, (T_{uv}$ when the LCA of u and v is in level $i + 1)$

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} pr(S_{iw} \wedge X_{iw}) 2^{i+3}$ (respectively denoted the above two descriptions with S_{iw} and X_{iw})

  $= \sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} pr(S_{iw} | X_{iw}) pr(X_{iw}) 2^{i+3}$

from here on, w.l.o.g., we suppose u is nearer to w than v

Upper bounding the expected distortion: $pr(S_{iw}|X_{iw})$

- $pr(S_{iw}|X_{iw}) \leq \frac{1}{j}$ if w is the j^{th} closest vertex to the pair u, v

Upper bounding the expected distortion: $pr(S_{iw}|X_{iw})$

- $pr(S_{iw}|X_{iw}) \leq \frac{1}{j}$ if w is the j^{th} closest vertex to the pair u, v
- therefore, $\sum_{w \in V} \sum_{i=0}^{\lg \Delta - 1} pr(S_{iw}|X_{iw}) pr(X_{iw}) 2^{i+3}$
  $= \sum_{j=1}^{n} \frac{1}{j} \sum_{i=0}^{\lg \Delta - 1} pr(X_{iw}) 2^{i+3}$ (since for each j, $1 \leq j \leq n$, there is some vertex w that is the j^{th} closest to the pair u, v)

Upper bounding the expected distortion: $pr(X_{iw})$

- $pr(X_{iw}) = pr(u \in B(w, r_i) \text{ and } v \notin B(w, r_i))$
  $= \frac{|[2^{i-1}, 2^i) \cap [d_{uw}, d_{vw})|}{|[2^{i-1}, 2^i)|}$ (since $r_i \in [2^{i-1}, 2^i)$)
  $= \frac{|[2^{i-1}, 2^i) \cap [d_{uw}, d_{vw})|}{2^{i-1}}$

Upper bounding the expected distortion: $pr(X_{iw})$

- $pr(X_{iw}) = pr(u \in B(w, r_i) \text{ and } v \notin B(w, r_i))$
  $= \frac{|[2^{i-1}, 2^i) \cap [d_{uw}, d_{vw})|}{|[2^{i-1}, 2^i)|}$ (since $r_i \in [2^{i-1}, 2^i)$)
  $= \frac{|[2^{i-1}, 2^i) \cap [d_{uw}, d_{vw})|}{2^{i-1}}$

- $\sum_{i=0}^{\lg_2 \Delta - 1} 2^{i+3} pr(X_{iw})$
  $= 16 \sum_{i=0}^{\lg_2 \Delta - 1} |[2^{i-1}, 2^i) \cap [d_{uw}, d_{vw})|$
  $\leq 16 |[d_{uw}, d_{vw})|$ (since the intervals $[2^{i-1}, 2^i)$ for $i = 0$ to $\lg_2 \Delta - 1$ partition the interval $[1/2, \Delta/2)$)
  $= 16(d_{vw} - d_{uw})$
  $\leq 16 d_{uv}$

Upper bounding the expected distortion: $E[T_{uv}]$

- $E[T_{uv}]$

  $= \sum_{j=1}^{n} \frac{1}{j} \sum_{i=0}^{\lg \Delta - 1} pr(X_{iw}) 2^{i+3}$

  $\leq \sum_{j=1}^{n} \frac{1}{j} (16 d_{uv})$

  $= 16 d_{uv} \sum_{j=1}^{n} \frac{1}{j}$

hence, $E[T_{uv}]$ is $O(\lg n) d_{uv}$

Outline

Transforming Tree Metric (V', T) to a Spanning Tree Metric (V, T')

1. repeat until there does not exist a vertex pair u, w such that $u \in V$, $w \notin V$ and w is the parent of u
   (a) contract edge uw
   (b) identify merged node with $u (\in V)$
2. multiply the length of every remaining edge by four

Distortion in Spanning Tree Metric

- $T'(u, v) \leq 4T(u, v)$ for any $u, v \in V$

  immediate from the construction

Distortion in Spanning Tree Metric

- $T'(u,v) \leq 4T(u,v)$ for any $u,v \in V$

  immediate from the construction

- $T(u,v) \leq T'(u,v)$ for any $u,v \in V$
  - if LCA of u and v is in T was a node w at level i so that $T_{uv} = 2^{i+2} - 4$
  - the contraction process only moves u and v upward in T, the distance T'_{uv} must be at least 4 times the length of the edge from w to one of its children

Distortion in Spanning Tree Metric

- $T'(u, v) \leq 4T(u, v)$ for any $u, v \in V$

  immediate from the construction

- $T(u, v) \leq T'(u, v)$ for any $u, v \in V$
  - if LCA of u and v is in T was a node w at level i so that $T_{uv} = 2^{i+2} - 4$
  - the contraction process only moves u and v upward in T, the distance T'_{uv} must be at least 4 times the length of the edge from w to one of its children

hence, $d_{uv} \leq T'_{uv}$ and $E[T'_{uv}]$ is $O(\lg n)d_{uv}$

Outline

Metric k-median clustering problem

Given a set P of n points in metric space, find a set $S \subseteq P$ of $k > 0$ points (a.k.a., cluster centers), such that the sum of the distances of points of P to their closest point in S is minimized.

$$\min_{S \subseteq P, |S| = k} \sum_{p \in P} dist(p, S)$$

(hence, a.k.a. min-sum clustering)

Algorithm

1. embed the finite metric space (P, d) into a tree metric (T, d_T)
2. convert T into a binary tree; optimally solve k-median clustering problem on the resultant tree with points at the leaves
3. output the so obtained set C of k-centers together with the corresponding clusters

DP to solve k-median clustering problem on a tree with points at leaves

- observation: for the leaf nodes $1, \ldots, i, \ldots, j, \ldots, r, \ldots, n$ of the tree metric, $d(i,j) \leq d(i,k)$

DP to solve k-median clustering problem on a tree with points at leaves

- observation: for the leaf nodes $1, \ldots, i, \ldots, j, \ldots, r, \ldots, n$ of the tree metric, $d(i,j) \leq d(i,k)$

- let $opt_r(i,j)$ denote the optimal solution with r-centers and points at leaves $l_i, \ldots, l_j$

  then, $opt_k(1,n) =$

  $= \min_{1 \leq j \leq n-k}(opt_1(1,j) + opt_{k-1}(j+1,n))$

— $O(k^2 n)$ time

Expected Approximation

- $cost_C(P, d)$
  $\leq cost_C(P, d_T)$
  $\leq cost_{C_{opt}}(P, d_T)$
  $= \sum_{p \in P} d_T(p, C_{opt})$
  $\leq \sum_{p \in P} d_T(p, \text{center associated to } p \text{ in } C_{opt})$

Expected Approximation

- $cost_C(P, d)$
  $\leq cost_C(P, d_T)$
  $\leq cost_{C_{opt}}(P, d_T)$
  $= \sum_{p \in P} d_T(p, C_{opt})$
  $\leq \sum_{p \in P} d_T(p, \text{center associated to } p \text{ in } C_{opt})$
- hence, $E[cost_C(P, d)]$ is $O(OPT \cdot \lg n)$

A few more applications

- Uniform buy-at-bulk network design
- Group Steiner tree
- Vehicle routing
- Communication spanning trees

Outline

Other popular metric embeddings

Let $|X| = n$, d' is an arbitrary metric, d denotes the number of dimensions, and X' is finite. Then,

- Bourgain's theorem:

  existence of $(X, d') \hookrightarrow^{O(\lg n)} (\mathcal{R}^{O(\lg^2 n)}, L_p)$

Other popular metric embeddings

Let $|X| = n$, d' is an arbitrary metric, d denotes the number of dimensions, and X' is finite. Then,

- Bourgain's theorem:

  existence of $(X, d') \hookrightarrow^{O(\lg n)} (\mathcal{R}^{O(\lg^2 n)}, L_p)$

- dimension reduction due to the Johnson-Lindenstrauss lemma:

  existence of $(X, L_2^d) \hookrightarrow^{(1+\epsilon)} (\mathcal{R}^{O(\epsilon^{-2} \lg n)}, L_2)$

Other popular metric embeddings

Let $|X| = n$, d' is an arbitrary metric, d denotes the number of dimensions, and X' is finite. Then,

- Bourgain's theorem:

  existence of $(X, d') \hookrightarrow^{O(\lg n)} (\mathcal{R}^{O(\lg^2 n)}, L_p)$

- dimension reduction due to the Johnson-Lindenstrauss lemma:

  existence of $(X, L_2^d) \hookrightarrow^{(1+\epsilon)} (\mathcal{R}^{O(\epsilon^{-2} \lg n)}, L_2)$

- Feige's volume respecting embeddings:

  $Vol(X) = sup_{f:X \to l_2} Evol(f(X))$ (f requires to be a contraction)

  k-distortion of f is $sup_{P \subset X, |P|=k} \left(\frac{Vol(P)}{Evol(f(P))} \right)^{\frac{1}{k-1}}$

Other popular metric embeddings

Let $|X| = n$, d' is an arbitrary metric, d denotes the number of dimensions, and X' is finite. Then,

- Bourgain's theorem:

  existence of $(X, d') \hookrightarrow^{O(\lg n)} (\mathcal{R}^{O(\lg^2 n)}, L_p)$

- dimension reduction due to the Johnson-Lindenstrauss lemma:

  existence of $(X, L_2^d) \hookrightarrow^{(1+\epsilon)} (\mathcal{R}^{O(\epsilon^{-2} \lg n)}, L_2)$

- Feige's volume respecting embeddings:

  $Vol(X) = sup_{f:X \to l_2} Evol(f(X))$ (f requires to be a contraction)

  k-distortion of f is $sup_{P \subset X, |P|=k} \left(\frac{Vol(P)}{Evol(f(P))}\right)^{\frac{1}{k-1}}$

* It is known that $\Omega(\lg n)$ distortion is necessary in the worst-case.

Possible future work

- improving the apprx factor in case of shortest path metrics from special graphs
- several metric embedding conjectures listed in [Indyk01]
- simpler algorithms for non-uniform buy-at-bulk
- finding more applications for tree metrics

References

- Y. Bartal. Probabilistic Approximation of Metric Spaces and its Algorithmic Applications. FOCS 1996, 184-193.
- J. Fakcharoenphol, S. Rao and K. Talwar A tight bound on approximating arbitrary metrics by tree metrics. Journal of Computer and System Sciences, 69:485-497, 2004.
- P. Indyk. Algorithms applications of low-distortion geometric embeddings. FOCS 2001.
- David P. Williamson and David B. Shmoys. The Design of Approximation Algorithms. Cambridge University Press, 2011.
- Jiri Matousek. Lectures on Discrete Geometry. Springer, 2002.
- Sariel Har-Peled. Geometric Approximation Algorithms. Springer, 2011.

Thanks!