Randomized Algorithms 2017A - Lecture 10
Metric Embeddings into Random Trees

Lior Kamma

1 Introduction

Embeddings and Distortion. An embedding of a metric space (X, d_X) into a metric space (Y, d_Y) is a map $f : X \to Y$. Its (bi-Lipschitz) distortion is the least $D \geq 1$ such that

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

Some related results previously seen in class

Claim. *Every n-point metric space embeds isometrically (i.e., with distortion 1) into ℓ_∞^n.*

Theorem (Bourgain 1985)**.** *Every n-point metric space embeds into ℓ_2 with distortion $O(\log n)$.*

Theorem (Johnson-Lindenstrauss)**.** *Every n-point metric subspace of ℓ_2^d embeds into ℓ_2^k with distortion $(1 + \varepsilon)$, where $k = O(\varepsilon^{-2} \log n)$.*

Tree Metrics. Consider an undirected graph $G = (V, E)$ with non-negative edge weights $\{w_e\}_{e \in E}$.

Exercise: Show that the function $d_G : V \times V \to \mathbb{R}$, which maps every pair $x, y \in V$ to the length of a shortest path between x and y in G w.r.t. w, is a metric on V.

A metric space (Y, d_Y) is called a *tree metric space* if there exists a tree G such that Y embeds isometrically into G.

"Dream Goal": Embed an arbitrary metric space (X, d_X) into a tree metric space with "small" distortion.

Motivation. We first note that every finite tree metric space can be embedded isometrically into ℓ_1.

Exercise: Prove it.

Additionally, many optimization and online problems involve a metric defined on a set of points. It is often useful to embed a metric space into a simpler one while keeping the distances approximately. Specifically, many such problems can be efficiently solved or better approximated on trees.

Bear the following example, called *k-median*, in mind. You are given a metric space (X, d_X) and an integer k. The goal is to choose a set $S \subseteq X$ of size at most k, that minimizes the objective function $\sum_{x \in X} d_X(x, S)$. This problem is known to be NP-Hard, however it can be solved optimally on trees in polynomial time. The heuristic is as follows. Embed X into a tree metric Y, solve the problem on Y, and construct a respective solution in X.

Details are omitted at this point, mainly due to the fact that, unfortunately, this approach does not work so well.

Embedding a Cycle into a Single Tree. Let (C_n, d_{C_n}) denote the shortest-path metric on an unweighted n-cycle. One can easily show show that embedding the cycle into a spanning tree incurs a distortion $D \geq \Omega(n)$. In fact, Rabinovich and Raz [RR98] showed that every embedding of the cycle into a tree (not necessarily a spanning tree, and may have additional vertices) incurs distortion $\geq \Omega(n)$.

2 Randomized Embeddings

However, not all is lost. If we consider a *random* embedding of C_n, then we can bound the distortion *in expectation.* Let T be the random tree that results from deleting a single edge of C_n chosen uniformly at random. Notice that this embedding satisfies the following two properties (proved in class).

1. For every $x, y \in C_n$. $d_{C_n}(x, y) \leq d_T(x, y)$.
2. For every $x, y \in C_n$. $\mathbb{E}[d_T(x, y)] \leq 2d_{C_n}(x, y)$.

Exercise: Extend the result to a weighted cycle.

New Goal. Embed an arbitrary metric space (X, d_X) into a random dominating tree metric with "small" *expected* distortion.

In fact, we will show a somewhat stronger result.

Definition 1. *A* k-hierarchically well-separated tree (k-HST) *is a rooted weighted tree* $T = (V(T), E(T))$ *satisfying the following properties.*

1. *For every node* $v \in V(T)$*, all edges connecting* v *to a child are of equal weight.*
2. *The edge weight along a path from the root to a leaf decrease by a factor of at least* k*.*

Theorem 1 ([FRT04])**.** *Let (X, d_X) be an n-point metric space. There exists a randomized polynomial-time algorithm that embeds X into the set of leaves of a 2-HST $T = (V(T), E(T))$ such that the following holds (we may assume that $X \subseteq V(T)$).*

1. *For every $x, y \in X$. $d(x, y) \leq d_T(x, y)$.*
2. *For every $x, y \in X$. $\mathbb{E}[d_T(x, y)] \leq O(\log n) d_X(x, y)$.*

Note that since the distortion is bounded in expectation, we can still apply the approximation heuristic considered earlier for problems in which the objective function is linear.

Back to k-Median.

Lemma 1. *The k-median problem can be solved efficiently on the metric space induced by the set of leaves of a 2-HST.*

Exercise: Prove Lemma 1. **Hint:** Use dynamic programming.

Corollary 1. *There exists a randomized approximation algorithm for the k-median problem with expected ratio $O(\log n)$.*

proof sketch. Given a metric space (X, d_X) and an integer k, we apply Theorem 1 and randomly embed X into a 2-HST T. We solve the problem on the leaves of T and return the solution. □

3 Partitions, Laminar Families and Trees.

Definition 2. *A set-family $\mathcal{L} \subseteq 2^X$ is called* laminar *if for every $A, B \in \mathcal{L}$, if $A \cap B \neq \emptyset$ then $A \subseteq B$ or $B \subseteq A$.*

A laminar family $\mathcal{L} \subseteq 2^X$ such that $\{x\} \in \mathcal{L}$ for all $x \in X$, induces a tree T such that $V(T) = \mathcal{L}$ and the leaves of T are exactly $\{\{x\} : x \in X\}$ in a straightforward manner.

We can construct a laminar family by repeatedly partitioning X. In order to make sure the algorithm halts, we can, e.g. decrease the diameter of the sets in the partition in each iteration. Let Π be a partition of X. Every $A \in \Pi$ is called a *cluster*, and for every $x \in X$, let $\Pi(x)$ denote the unique cluster $A \in \Pi$ such that $x \in S$. Denote the diameter of X by Δ. By scaling we may assume without loss of generality that $\min_{x,y \in X} d_X(x, y) = 1$.

Input: X.
Output: A laminar family $\mathcal{L} \subseteq 2^X$ such that $\{\{x\} : x \in X\} \subseteq \mathcal{L}$.

$\Pi_0 \leftarrow \{X\}$, $\mathcal{L} \leftarrow \{X\}$
for $i = 1$ to $\log \Delta$ **do**
  $\Pi_i \leftarrow \emptyset$.
  for all $A \in \Pi_{i-1}$ **do**
    if $|A| > 1$ **then**
      Let Π be a partition of A into clusters of diameter at most $2^{-i}\Delta$.
      $\Pi_i \leftarrow \Pi_i \cup \Pi$.
      $\mathcal{L} \leftarrow \mathcal{L} \cup \Pi$.
return $\mathcal{L}$.

Algorithm 1: Constructing a Laminar Family

It remains to show how to construct the partitions Π_i, $i \in [\log \Delta]$, and how to set the weights of the tree edges.

4 From Low-Diameter Decompositions to Low-Distortion Embeddings

Definition 3. *A metric space (X, d_X) is called β-*decomposable *for $\beta > 0$ if for every $\delta > 0$ there is a probability distribution μ over partitions of X, satisfying the following properties.*

(a). Diameter Bound: For every $\Pi \in \mathsf{supp}(\mu)$ and $A \in \Pi$, $diam(A) \leq \delta$.

(b). Separation: For every $x, y \in X$,

$$\Pr_{\Pi \sim \mu}[\Pi(x) \neq \Pi(y)\ ] \leq \beta \cdot \frac{d_X(x,y)}{\delta}\ .$$

Theorem 2 ([Bar96], [FRT04])**.** *Every n-point metric space is $8 \log n$-decomposable.*

In fact, Fakcharoenphol, Rao and Talwar [FRT04] gave a somewhat stronger result, which will prove essential in the analysis of the embedding. We replace the separation property in Definition 3 by the following, stronger requirement.

(b'). For every $x, y \in X$, if $d_X(x,y) < \frac{\delta}{8}$ then

$$\Pr_{\Pi \sim \mu}[\Pi(x) \neq \Pi(y)\ ] \leq \frac{d_X(x,y)}{\delta} \cdot 8 \log \frac{|B(\{x,y\}, \delta/2)|}{|B(\{x,y\}, \delta/8)|}\ ,$$

where $B(\{x,y\}, r) = \{z \in X : d_X(\{x,y\}, z) \leq r\}$ for all $r > 0$.

We can now update Algorithm 1 and construct the tree embedding.

Input: X.
Output: A 2-HST with X being the set of leaves.

```
Π_0 ← {X}.
V(T) ← Π_0, E(T) ← ∅.
for i = 1 to log Δ do
   Π_i ← ∅.
   for all A ∈ Π_{i−1} do
      if |A| > 1 then
         Let Π be a random partition of A as in Theorem 2 with δ = 2^{−i}Δ.
         Π_i ← Π_i ∪ Π.
         V(T) ← V(T) ∪ Π.
         Add to E(T) an edge from A to every cluster in Π, of weight δ.
return T.
```

Algorithm 2: Constructing a Random Embedding into a 2-HST

Applying Theorem 2, we now turn to prove Theorem 1. Note that the leaves of T are exactly the sets $\{x\}$ for all $x \in X$, and thus for every $x \in X$, we can identify $\{x\} \in V(T)$ with x. Clearly T is a 2-HST. Consider next $x, y \in X$ and let i_0 be the unique integer such that $d_X(x, y) \in (2^{-i_0}\Delta, 2^{-(i_0-1)}\Delta]$, and let i^* be the first index for which $\Pi_{i^*}(x) \neq \Pi_{i^*}(y)$. By the diameter bound of the partition we get that $i^* \leq i_0$. We therefore conclude the following.

Claim 1. $d_T(x, y) \geq d_X(x, y)$.

Proof. $d_T(x, y) \geq 2 \cdot 2^{-i^*} \geq 2^{-i_0+1}\Delta \geq d_X(x, y)$. □

Claim 2. $d_T(x, y) \leq 2^{-i^*+2}\Delta$.

Proof. Denote by $u \in V(T)$ the least common ancestor of x, y. Consider the path from u to x. Since T is a 2-HST we get that the length of the path is at most

$$\sum_{i=i^*}^{\infty} 2^{-i}\Delta = 2^{-i^*+1}\Delta \ .$$

The length of the xy-path in T is at most twice as long. □

The following claim concludes the proof of Theorem 1.

Corollary 2. $\mathbb{E}[d_T(x, y)] \leq O(\log n) d_X(x, y)$.

Proof. Since $1 \leq i^* \leq i_0$, then $\mathbb{E}[d_T(x, y)] = \sum_{i=1}^{i_0} \mathbb{E}[d_T(x, y)|i^* = i] \cdot \Pr[i^* = i]$. By Claim 2, $\mathbb{E}[d_T(x, y)|i^* = i] \leq 2^{-i+2}\Delta$, and by Theorem 2, $\Pr[i^* = i] \leq \frac{d_X(x,y)}{2^{-i}\Delta} \cdot \log \frac{|B(\{x,y\},2^{-i}\Delta/2)|}{|B(\{x,y\},2^{-i}\Delta/8)|}$. Therefore

$$\mathbb{E}[d_T(x, y)] \leq \sum_{i=1}^{i_0} 2^{-i+2}\Delta \cdot \frac{d_X(x, y)}{2^{-i}\Delta} \cdot \log \frac{|B(\{x, y\}, 2^{-i-1}\Delta)|}{|B(\{x, y\}, 2^{-i-3}\Delta)|} = 4d_X(x, y) \sum_{i=1}^{i_0} \log \frac{|B(\{x, y\}, 2^{-i-1}\Delta)|}{|B(\{x, y\}, 2^{-i-3}\Delta)|} \ .$$

All but a constant number of elements of the sum are canceled, and therefore $\mathbb{E}[d_T(x,y)] \leq O(\log n) d_X(x,y)$. $\square$

5 Randomized Low-Diameter Decompositions

We now turn to prove Theorem 2. The following algorithm samples a partition of X. We will show that the distribution induced by the algorithm satisfies the conditions of the theorem.

Input: X, δ.
Output: A partition Π as in Theorem 2

$\Pi \leftarrow \emptyset$.
let π be a random ordering of X.
independently choose $R \in (\delta/4, \delta/2]$ uniformly at random.
for all $j \in [n]$ **do**
  let $B_j = B(\pi(j), R)$.
  let $C_j = B_j \setminus \bigcup_{j'<j} B_{j'}$
  if $C_j \neq \emptyset$ then $\Pi \leftarrow \Pi \cup \{C_j\}$.
return Π.

Algorithm 3: Constructing a Random Partition

Clearly for every $C \in \Pi$, $\mathsf{diam}(C) \leq \delta$. Fix $x, y \in X$, and let $x_1, x_2, \ldots, x_n$ be an ordering of X in ascending distance from $\{x,y\}$ (breaking ties arbitrarily). Fix $j \in [n]$. We say that x_j *settles* x, y if $B(x_j, R)$ is the first ball (in the order induced by π) that has non-empty intersection with $\{x,y\}$. We say that x_j *cuts* x, y if $|B(x_j,R) \cap \{x,y\}| = 1$, and x_j *separates* x, y if x_j both settles x, y and cuts x, y.

Notice that the event that x_j cuts x, y depends only on the choice of R and is independent of the choice of π. Assume, without loss of generality that $d_X(j,x) \leq d_X(j,y)$.

Claim 3. $\Pr[x_j \textit{ separates } x, y] \leq \frac{1}{j} \cdot \frac{4d_X(x,y)}{\delta}$.

Proof. First note that

$$\Pr[x_j \text{ separates } x,y] = \Pr[x_j \text{ separates } x,y \mid x_j \text{ cuts } x,y] \cdot \Pr[x_j \text{ cuts } x,y] \,.$$

Note that x_j cuts x, y if and only if $R \in [d_X(j,x), d_X(j,y))$. Since R is uniformly distributed over $(\delta/4, \delta/2]$, and from the triangle inequality we get that

$$\Pr[x_j \text{ cuts } x,y] = \Pr[R \in [d_X(j,x), d_X(j,y))\ ] \leq \frac{d_X(j,y) - d_X(j,x)}{\delta/4} \leq \frac{4d_X(x,y)}{\delta} \,.$$

Conditioned on x_j cutting x, y, assume toward contradiction that there exists $j' < j$ such that $x_{j'}$ precedes x_j in the order induced by π. Since $d_X(x_{j'}, \{x,y\}) \leq d_X(x_j, \{x,y\}) =$

$d_X(x_j, x) \leq R$, it follows that $\{x, y\} \cap B(x_{j'}, R) \neq \emptyset$ and therefore x_j does not settle x, y, a contradiction. Therefore,

$$\Pr[x_j \text{ separates } x, y \mid x_j \text{ cuts } x, y] \leq \Pr[x_j \text{ precedes } x_{j'} \text{ for all } j' < j] \leq \frac{1}{j}$$

□

Since $\Pr[\Pi(x) \neq \Pi(y)] \leq \sum_{j \in [n]} \Pr[x_j \text{ separates } x, y]$ we get that

$$\Pr[\Pi(x) \neq \Pi(y)] \leq \sum_{j \in [n]} \frac{4 d_X(x, y)}{j\delta} \leq 4 \frac{d_X(x, y)}{\delta} \cdot (\log n + 1) \leq \frac{d_X(x, y)}{\delta} \cdot 8 \log n \ .$$

In order to get a stronger result, we need a more delicate analysis. Assume that $d_X(x, y) \leq \delta/8$, then if $x_j \in B(\{x, y\}, \delta/8)$, then $\Pr[x_j \textit{ separates } x, y] = 0$. In addition, if $x_j \notin B(\{x, y\}, \delta/2)$, then $\Pr[x_j \textit{ separates } x, y] = 0$. Therefore

$$\Pr[\Pi(x) \neq \Pi(y)] \leq \sum_{j \in B(\{x,y\},\delta/2) \setminus B(\{x,y\},\delta/8)} \frac{4 d_X(x, y)}{j\delta} \leq \frac{d_X(x, y)}{\delta} \cdot 4 \log \frac{|B(\{x, y\}, \delta/2)|}{|B(\{x, y\}, \delta/8)|} \ .$$

Exercise: A metric space (X, d_X) is called *β-padded-decomposable* for $\beta > 0$ if for every $\delta > 0$ there is a probability distribution μ over partitions of X, satisfying the following properties.

(a). Diameter Bound: For every $\Pi \in \mathsf{supp}(\mu)$ and $A \in \Pi$, $diam(A) \leq \delta$.

(b). Padding: For every $x \in X$ and $\varepsilon < \delta/8$,

$$\Pr_{\Pi \sim \mu}[B(x, \varepsilon) \not\subseteq \Pi(x)] \leq \beta \cdot \frac{\varepsilon}{\delta} \ .$$

Show that every n-point metric space is $O(\log n)$-padded-decomposable.

References

[Bar96] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science*, pages 184–193. IEEE, 1996.

[FRT04] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, 2004.

[RR98] Y. Rabinovich and R. Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete Comput. Geom.*, 19(1):79–94, 1998.