Metric Techniques for Approximation Algorithms **Exercises 1**
Date: August 18, 2008

Notes: As usual, if $\mathsf{M} = (V, d)$ and $\mathsf{M}' = (V', d')$ are two metric spaces, then $\mathsf{M}_1 \overset{D}{\hookrightarrow} \mathsf{M}_2$ implies that there exists a map $f : V \to V'$ with distortion at most D. When we write $\mathsf{M}_1 \overset{\geq D}{\hookrightarrow} \mathsf{M}_2$, this is a *lower bound* statement: *every map* $f : V \to V'$ has distortion at least D.

This naturally extends to the case when $\mathcal{G}$ is a *family of metrics or graphs*, then $\mathsf{M} \overset{D}{\hookrightarrow} \mathcal{G}$ implies that *there exists* $\mathsf{M}' \in \mathcal{G}$ such that $\mathsf{M} \overset{D}{\hookrightarrow} \mathsf{M}'$; similarly, $\mathsf{M} \overset{\geq D}{\hookrightarrow} \mathcal{G}$ implies that *for all* $\mathsf{M}' \in \mathcal{G}$, it holds that $\mathsf{M} \overset{D}{\hookrightarrow} \mathsf{M}'$

---

1. Tree Metrics and a 4-point Inequality. A result of Bandelt says that a metric (V, d) is a tree-metric, if for any four points $i, j, k, l \in V$,

$$d_{ij} + d_{kl} \leq \max\{d_{ik} + d_{jl}, d_{il} + d_{jk}\}$$

(where we write $d(i, j)$ as d_{ij}, etc.) Infer that the two largest of the three sums $d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}$ are equal. Use this to give another proof that the shortest-path metric of C_4 is not a tree metric.

2. Embedding a Metric into a (Single) Tree. In this exercise, we investigate the distortion incurred when embedding graphs into (single) trees.

a. Show that any metric $\mathsf{M} = (V, d)$ can be embedded into a tree with distortion at most $n - 1$.

b. Show that any embedding of the (unweighted) n-cycle C_n into 1-dimensional space $\mathbb{R}^1$ incurs a distortion of at least $n - 1$. Equivalently, show that any embedding of C_n into a (weighted) path P must incur a distortion of at least $n - 1$.

c. Show that given any embedding of $C_n = (V_n, E_n)$ into *any tree* $T = (V_n, E)$—where the tree has exactly the same node set as the cycle—with distortion D, there exists a path P such that $C_n \overset{D}{\hookrightarrow} P$. Infer that $C_n \overset{\geq (n-1)}{\hookrightarrow}$ *the family of trees on* n *nodes.*

d. Give an example to show that there are trees $T = (V, E)$ such that $V_n \subsetneq V$ where $C_n \overset{< n-1}{\hookrightarrow} T$: that is, having Steiner nodes *does* help. (Hint: think small cycles.)

e. However, it is still known that $C_n \overset{\Omega(n)}{\hookrightarrow}$ trees. A result of Rabinovich and Raz (1998) proves that any embedding of the n-cycle into a tree (regardless of number of Steiner nodes) incurs a distortion of $\frac{n}{3} - 1$. (So, Steiner nodes don't help *too much.*)

We'll prove a weaker result here. Gupta (2000) shows that given any (weighted) tree $T = (X \cup Y, E)$, there is a (weighted) tree $T_X = (X, E_X)$ only on the nodes in X such that for any $x, x' \in X$, $1 \leq \frac{d_T(x,x')}{d_{T_X}(x,x')} \leq 8$. Use this result, and the above parts, to show a lower bound of $> \frac{n-1}{8}$ for embedding cycles into trees.

f. Finally, show that the $\sqrt{n} \times \sqrt{n}$-square grid greph $\square_n$ incurs a distortion of $\Omega(\sqrt{n})$ when embedded into trees. (Can you find a subset of $t = \Theta(\sqrt{n})$ points in $\square_n$ which embed into a cycle C_t with constant distortion?)

Open question: is it possible to somehow use the 4-point results from Problem 1 to give the lower bounds proved in this problem?

3. Tree Embeddings and Approximation Algorithms. Given a metric space $\mathsf{M} = (V, d)$ on $|V| = n$ points, let $\mathcal{T}$ be the set of trees $T = (V, E_T)$ with edge lengths $\ell : E_T \to \mathbb{R}$ such that for each $u, v \in V$, $d_T(u, v) \geq d(u, v)$—i.e., these are trees whose distances *dominate* those in M. A probability distribution $\mathcal{D}$ on this set of "dominating" trees $\mathcal{T}$ is said to α-*approximate* the metric $\mathcal{M}$ if for every $u, v \in V$,

$$\mathbf{E}_{T \leftarrow \mathcal{D}}[d_T(u, v)] \leq \alpha \cdot d(u, v). \tag{1}$$

We saw the *tree-embedding* results of Bartal (*FOCS 1996*) and Fakcharoenphol, Rao, & Talwar (*STOC 2003*) that given any metric M on n points, one can find a distribution $\mathcal{D} = \mathcal{D}(\mathsf{M})$ such that $\alpha = \alpha_{FRT} = O(\log n)$. In fact, this embedding is efficient: in poly-time one can output a random tree from this distribution π.

In this exercise, we give details on how to use this result to design approximation algorithms for problems defined on general metric spaces by first designing exact or near-optimal solutions to the same problem on trees, and then using the Bartal/FRT results to translate such an algorithm to arbitrary metrics with a loss of $\alpha = O(\log n)$ in the performance ratio.

- **a**. Consider the traveling salesman problem (TSP): given an n-point metric (X, d), find a permutation of the points to minimize the TSP tour induced by this permutation. Suppose you are given a tree T with nonnegative edge lengths as your TSP input, show how to solve the TSP problem in polynomial time for this class of input graphs.

- **b**. To solve TSP on a general metric, we apply above embedding result to the input metric to derive a probability distribution over tree metrics. We then sample a tree from this distribution and apply your polynomial time algorithm (from the previous part) to it, and output the resulting permutation. Show that this is an α-approximation to the TSP on the original metric.

- **c**. Where in your proof do you use the fact that distances in the random tree T dominate those in d_G (with probability 1)? If you just had the property that

  $$d(u, v) \leq \mathbf{E}_{T \leftarrow \mathcal{D}}[d_T(u, v)] \leq \alpha \cdot d(u, v),$$

  would your proof still go through?

n.b. It is well-known how to get a 1.5-approximation for TSP, and the $O(\log n)$ approximation given here is just for ease of explanation. However, there are several problems for which the best approximation known is still via such a randomized reduction to trees.

4. An Optimal Embedding into Trees. Recall the padded decomposition procedure given in the notes: Given a metric $\mathsf{M} = (V, d)$,

1. Pick a random permutation π on V.
2. Pick a random radius R uniformly from the interval $(\Delta/4, \Delta/2]$.
3. Create a "cluster" C_v for each $v \in V$: assign $x \in V$ to C_v
   if v is the *first* vertex (according to π) such that $d(v, x) \leq R$.
4. Output all the non-empty clusters C_v.

Clearly, all the clusters had diameter $\leq \Delta$; moreover, we saw that for any $x \in V$,

$$\Pr[\mathbf{B}(x, \rho) \text{ split by partitioning }] \leq \beta \frac{\rho}{\Delta}$$

for $\beta = O(\log n)$.

Recap of the proof: we named the vertices in V as $v_1, v_2, \ldots, v_n$ in increasing distance from x, and defined v_j as "responsible for splitting $B(x, \rho)$" if (i) the random value $R \in [d(x, v_j) - \rho, d(x, v_j) + \rho]$ *and* (ii) the node v_j came before $v_1, v_2, \ldots, v_{j-1}$ in the random order π. We claimed that if $B(x, \rho)$ was split, then at least one of the nodes was responsible for the split, and hence

$$\Pr[\mathbf{B}(x, \rho) \text{ split }] \leq \sum_j \Pr[v_j \text{ responsible }]$$

Since the two requirements (i) and (ii) depend on independent random choices, we got that $\Pr[v_j \text{ responsible }] = \frac{2\rho}{\Delta/4} \times \frac{1}{j}$. Summing this up gave the bound.

a. We claimed (without proof) a stronger result, where

$$\beta = \beta(x, \Delta) = O\left(\log \frac{|\mathbf{B}(x, \Delta)|}{|\mathbf{B}(x, \Delta/8)|}\right).$$

Prove this claim.

b. Recall the randomized recursive tree construction procedure as in the lecture:

Procedure FRT(X, i) (Invariant: $X \subseteq V$, diameter$(X) \leq 2^i$.)

1. If $|V| = 1$, return X.
2. Use β-padded decomposition procedure on X with diameter bound 2^{i-1}
   to get random partition $X_1, X_2, \ldots, X_k$.
3. For each j, recursively call FRT$(X_j, i-1)$ to get tree T_j with root v_j.
4. For each $j \geq 2$, attach edges (r_1, r_j) of length 2^i to get connected tree T.
5. Return resulting tree T with root $r = r_1$.

Recall that we initiate the process with $X = V$ and $i = \lceil \log_2 \text{diameter}(V) \rceil$.

Show that using the above more sophisticated bound on β, and this same procedure, the random tree T created satisfies:

$$\mathbf{E}_T[d_T(x, y)] \leq O(\log n) \times d(x, y)$$

for any $x, y \in V$.

5. A Lower Bounding Technique. We will now give details on how to prove lower bounds on embeddings into distributions over trees; the same idea is used to give lower bounds on randomized algorithms and hence is of some interest.

a. For any real-valued $m \times n$-matrix M, show that

$$\min_j \max_i M_{ij} \geq \max_i \min_j M_{ij}.$$

Now suppose we consider probability distributions $\mathbf{p}$ and $\mathbf{q}$ over the rows and columns of the matrix M, then one can show that

$$\min_{J \leftarrow \mathbf{q}} \max_i \mathbf{E}[M_{iJ}] \geq \max_{I \leftarrow \mathbf{p}} \min_j \mathbf{E}[M_{Ij}]. \tag{2}$$

It is not hard to prove, but you can skip this for the time being. (This is the easy part of the von Neumann Minimax Theorem, which says that the above inequality is actually an equality.)

b. Given a metric $\mathsf{M} = (V, d)$, imagine it to be a weighted graph $\mathsf{M} = (V, E = \binom{V}{2})$ with edge lengths $w_{\{u,v\}} = d(u,v)$. Consider the family $\mathcal{T}$ of trees that dominate distances in M, and construct a $|E| \times |\mathcal{T}|$-matrix A which has one row for each edge, and one column for each tree in T. Given the i^{th} edge $\{u, v\}$ and j^{th} tree T_j, set

$$A_{ij} = \frac{d_{T_j}(u,v)}{d(u,v)};$$

this is the distortion incurred by edge $\{u, v\}$ when we use tree T_j.

Now use (2) and show that to derive a lower bound of β for the distortion incurred in embedding M (or equivalently, G) into any distribution over dominating trees, it suffices to do the following: Give a probability distribution $\mathbf{p}$ over the edge set E so that given any tree $T \in \mathcal{T}$, the expected distortion

$$\mathbf{E}_{e=\{u,v\} \leftarrow \mathbf{p}}[\frac{d_T(u,v)}{d(u,v)}]$$

is at least β. (Note that this last expectation was taken over the probability distribition on edges.)

c. You already saw that the cycle C_n embeds into distributions over (sub)trees with $\alpha = 2(1 - \frac{1}{n})$. Use the above technique to show a lower bound of $\frac{4}{3} - o(1)$ on the distortion of any such embedding of (unweighted) C_n into trees.

Useful fact: given the unweighted n-cycle C_n and any tree T such that $d_T \geq d_{C_n}$—that is, T dominates distances in C_n, there is an edge $\{u, v\}$ of C_n such that the tree distance $d_T(u,v) \geq \frac{n}{3} - 1$.

d. Consider the diamond graphs defined in the Handout: in the graph G_k, the distance between the left and right ends is 2^k, and the min-cut between them is also 2^k. To show a lower bound for embedding this graph into a distribution over trees, let us use the above framework, and set the edge-distribution to be uniform over the 4^k edges in this graph G_k.

Fix any tree T: since we have chosen the uniform distribution, the goal is now to show that the expected distance in the tree T between the endpoints of a random edge of G_k is $\Omega(k)$. Note there are 2^{k-1} edge-disjoint copies of $C_{2^{k+1}}$ in this graph, and use the above parts to show that there are 2^{k-1} edges which will suffer a distortion of at least $2^{k+1}/3 - 1$. Now, since there are 2^k edge-disjoint copies of C_{2^k} in the graph, there are 2^k edges which will suffer a distortion of at least $2^k/3 - 1$. Continuing this way, infer that the expected distance between the endpoints of a random edge is $\Omega(k)$. (Be careful about possible over-counting!)

6. Covers. Here are two definitions that are often used in routing and distributed computing. The procedures we've seen today give us simple constructions, for some setting of the parameters:

a. Given a metric $\mathsf{M} = (V, d)$, an (α, k)-*tree cover* is a collection of trees $\mathcal{T} = \{T_1, T_2, \ldots, T_k\}$ such that for any pair of nodes $x, y \in V$, there exists a tree $T_j \in \mathcal{T}$ such that

$$d(x, y) \leq d_{T_j}(x, y) \leq \alpha \cdot d(x, y).$$

Show that, given the FRT result showing a probability distribution $\mathcal{D}$ that α-approximates the metric M, one can obtain a $(2\alpha, O(\log n))$-tree cover from it via a randomized algorithm.

b. Given a metric $\mathsf{M} = (V, d)$, an (c, r, t)-*neigborhood cover* is a collection $\mathcal{S} = \{S_1, S_2, \ldots\}$ of subsets $S_i \subseteq V$ of points such that (a) for each point $x \in V$, there is a subset S_j that contains the r-ball $B(x, r) = \{x' \in V \mid d(x, x') \leq r\}$, (b) each point $x \in V$ is contained in at most t of the subsets in $\mathcal{S}$, and (c) each subset S_i has diameter at most cr.

Given a β-padded decomposition procedure, show how to get an $(2\beta, r, O(\log n))$-neighborhood cover for any $r > 0$.

7. Small Support Distributions. Given a metric M, if there exists a distribution $\mathcal{D}$ over trees that achieves a distortion of α, show that there *exists* a distribution on only $\binom{n}{2}$ trees that achieves the same distortion. (Hint: write an exponential-size linear program, and use elementary facts about vertex solutions of linear programs.)

Note: The construction of FRT from the lecture can be derandomized to give a family of polynomially many trees $\mathcal{T}_{FRT}$ such that the uniform distribution on this family also $O(\log n)$-approximates the metric M. Details can be found in the paper of Fakcharoenphol Rao and Talwar.

7. MultiCut. Recall that in multicut, you were given a graph $G = (V, E)$ and pairs $\{s_i, t_i\} \subseteq V$. The goal was to delete a minimum cardinality set of edges so that no s_i lies in the same connected component as t_i.

a. Show that the multicut problem even on star graphs is NP-hard by showing a reduction from the vertex cover problem.

b. Give a better padded decomposition for tree metrics, and use that to give a constant-factor approximation for the multicut problem on trees.