Thomas Rothvoß
Location: AAC 132 Discussion: 5.05.10

Exercises

Approximation Algorithms

Spring 2010

Sheet 9

Reminder: On May 5, lecture and tutorial are moved to AAC 132.

Exercise 1

Recall that for the k-TSP problem, we are given a complete graph $G = (V,E)$ with a metric cost function $c : E \to \mathbb{Q}_+$ and a parameter $k \in \{1,\dots,n\}$. The goal is to find a minimum length tour, visiting *at least k nodes*.

i) Show that if c is a tree metric (and you know the underlying tree T), then one can find an optimum tour in polynomial time.

ii) Give an expected $O(\log n)$-approximation algorithm for k-TSP in general metric graphs. Can you derandomize it?

Solution:

i) We show the claim by dynamic programming. The dynamic program will be easier to state for a slightly more general problem, where the set of nodes V is partitioned into *required vertices* R and *Steiner nodes* S and we have to visit at least k many different required nodes at least once (by short-cutting one can again obtain a tour that does not revisit a node and does not visit more than k nodes). Then it would not make sense to use an edge (u,v) that is not contained in the tree — instead we can buy all edges on the u-v path in T for the same price. In other words, it suffices to use only tree edges. We guess a node $r \in V$ that is visited by the optimum tour and consider r as the root of the tree.

The degrees in T might be arbitrary. To make our life easier, by inserting new Steiner nodes and cost 0 edges, we turn T into a tree with out-degree ≤ 2. Furthermore, by adding Steiner nodes, we may assume that required nodes have exactly $\{0,1\}$ children and Steiner nodes have degree 2 children.

For any node $v \in V$, and $k' \in \{0,\dots,k\}$ we define table entries

$$A(v,k') = \text{cheapest tour in the subtree below } v, \text{ starting and ending at } v \text{ that visits at least } k' \text{ required nodes}$$

If $v \in R$ and v has one child v_1, then

$$A(v,k) = A(v_1,k-1) + 2 \cdot d^T(v,v_1)$$

If $v \in S$ and has 2 children v_1,v_2, we use:

$$A(v,k) = \min\Big\{A(v_1,k) + 2\cdot d^T(v,v_1), A(v_2,k) + 2\cdot d^T(v,v_2), \min_{k=k_1+k_2}\Big\{A(v_1,k_1) + 2d^T(u,v_1) + A(v_2,k_2) + 2d^T(v,v_2)\Big\}\Big\}$$

ii) We use the theorem from the last slide of the tree embedding section to obtain trees $T_1,\dots,T_q$ with cost d_i, weight λ_i. Suppose we choose a a tree T from $T_1,\dots,T_q$ (i.e. $\Pr[T=T_i]=\lambda_i$), and let d^T be the induced tree metric. Then for any $u,v\in V$

- $c(u,v)\leq d^T(u,v)$
- $E[d^T(u,v)]\leq O(\log n)\cdot c(u,v)$

Let OPT^{T_i} be the cost of the optimum k-TSP solution w.r.t. metric d^T. We claim that $E[OPT^T]\leq O(\log n)\cdot OPT$. This can be easily seen as follows: Let E^* be the edges of the optimum tour in G. Then the same set is still a valid tour w.r.t. d^T. And

$$E[OPT^T]\leq E\Big[\sum_{(u,v)\in E^*} d^T(u,v)\Big] \overset{\text{linearity of expectation}}{=} \sum_{(u,v)\in E^*} E[d^T(u,v)]\leq O(\log n)\cdot c(u,v)$$

By i), we can compute a tour using edges E' of expected cost

$$E[\sum_{(u,v)\in E'} d^T(u,v)]\leq E[OPT^T]\leq O(\log n)\cdot OPT.$$

Then at least one of the trees T_i (with induced tree metric d^{T_i}) must have cost

$$\sum_{(u,v)\in E'} d^{T_i}(u,v)\leq E[OPT^T]\leq O(\log n)\cdot OPT$$

Since $d^{T_i}(u,v)\geq c(u,v)$, the tour E' is also not more expensive w.r.t. the original costs.

Exercise 2

For STEINER FOREST, the input is a complete, undirected graph $G=(V,E)$ with metric cost function $c:E\to\mathbb{Q}_+$ and pairs $(s_1,t_1),\dots,(s_k,t_k)$ ($s_i,t_i\in V$). The goal is to find a min cost subgraph H, that connects each s_i-t_i pair:

$$OPT=\min_{H\subseteq E}\Big\{\sum_{e\in H} c(e)\mid \forall i=1,\dots,k: H \text{ connects } s_i \text{ and } t_i\Big\}$$

(there is no need to connect s_i,t_j for $i\neq j$, hence H itself does not need to be connected. In fact, in general it will be a *forest*, that is a collection of trees). Consider the following linear program

$$\begin{aligned} \min \sum_{e\in E} x_e c_e & \\ \sum_{e\in\delta(S)} x_e &\geq 1 \quad \forall i=1,\dots,k\ \forall S\subseteq V: s_i\in S, t_i\notin S\\ x_e &\geq 0 \end{aligned}$$

Here x_e can be interpreted as a variable that indicates whether e is included in H or not. Prove that the integrality gap of this LP is upperbounded by $O(\log n)$.

Solution:

Let x_e^* be an optimum fractional solution of cost $OPT_f = \sum_{e\in E} x_e^* c_e$. Let T be a random tree and d^T be the induced tree embedding with $O(\log n)$-distortion (which exists using a theorem from the lecture). Then the cost of x^* in the dominating metric is

$$E[\sum_{e\in E} d^T(e)\cdot x_e^*] = \sum_{e\in E} x_e^* E[d^T(e)] \leq \sum_{e\in E} x_e^* O(\log n)\cdot c(e) \leq O(\log n)\cdot OPT_f$$

Next, for any edge $e=(u,v)\notin T$, we install x_e^* units on each edge on the u-v path in tree without that the cost increases. We obtain a new fractional solution y_e^* where $y_e^*=0$ if $e\notin T$ (in fact this is a feasible solution, since any cut S with $u\in S, v\notin S$ contains also at least one from any u-v path). Consider now any edge $e\in T$. Removing e from T gives 2 subtrees T_1, T_2 (i.e. $T = T_1\cup T_2\cup\{e\}$). If one has $s_i\in T_1$ and $t_i\in T_2$ for some i (or vice versa), then the cut constraint for $S := V(T_1)$ says that $y_e^*\geq 1$. Let us define

$$z_e^* := \begin{cases} 1 & \text{if } e \text{ separates an } s_i - t_i \text{ pair in } T \\ 0 & \text{otherwise} \end{cases},$$

which is a feasible and integral LP solution with $z_e^*\leq y_e^*$. Furthermore

$$E\left[\sum_{e\in E} d^T(e)\cdot z_e^*\right] \leq \sum_{e\in E} E[d^T(e)] y_e^* = \sum_{e\in E} E[d^T(e)]\cdot x_e^* = O(\log n)\cdot OPT_f$$

Of course, $H := \{e\in E \mid z_e^*=1\}$ is a feasible solution that is not more expensive w.r.t. the original cost c than $O(\log n)\cdot OPT_f$. By the principle of the probabilistic method, there must be one tree embedding and hence one concrete solution H, that really costs at most $O(\log n)\cdot OPT_f$. Hence the integrality gap is bounded by $O(\log n)$.
