Steiner forest analysis

Theorem 1 *The primal-dual algorithm presented in class is a 2-approximation for the Steiner forest problem.*

Let A denote the cost of the output. Let (x'_e) be the characteristic vector of the output forest F'. We have:

$$A = \sum_e c_e x'_e = \sum_e (\sum_{S:e\in\delta(S)} y_S) x'_e = \sum_S (\sum_e \in \delta(S) x'_e) y_S$$

because x'_e is only non-zero when x_e is non-zero, which can only happen when the corresponding dual constraint is tight. Let $\text{degree}_{F'}(S) = \sum_{e\in\delta(S)} x'_e$ be the number of edges of F' coming out of S.

Lemma 1 $\sum_S \textit{degree}_{F'}(S) y_S \leq 2\sum_S y_S$.

From the lemma, we obtain $A \leq 2\sum_S y_S \leq 2\text{OPT}$ since the fractional optimum is at most the integer optimum.

To prove the lemma, let us prove that $\Phi = 2\sum_S y_S - \sum_S \text{degree}_{F'}(S) y_S$ is non-negative. The proof is by induction as (y_S) changes over the execution of the algorithm. Initially $\Phi = 0$. At any time, if all currently active set variables y_S are increased by ϵ, the change in Φ is then

$$\epsilon(2\#(\text{active sets}) - \sum_{S \text{ active set}} \text{degree}_{F'}(S)).$$

We will prove that at any time during execution, this quantity is non-negative.

Draw the forest F just before the pruning leads to F'. Focus on a given time t during the execution. Mark the edges of F that have already been added (the edges of F such that $x_e = 1$ already at that time). Each active set is a connected component of marked edges, and each connected component of marked edges is either an active or an inactive set.

Modify F by contracting each connected component of marked edges to a node, defining a new forest F_t. In the contracted graph, remove the edges of F_t that are not in F', defining forest F'_t. Each active set S is a node of F'_t, and $\text{degree}_{F'}(S)$ is exactly the degree of that node in F'_t.

Without loss of generality we can assume that F'_t is a tree with at least one edge, since it is enough to prove that for each such tree $\sum_{u \text{ active}} \text{degree}_{F'}(u) \leq 2\#(u \text{ active})$ and sum.

I claim that every leaf of F'_t is active. To prove it, consider a leaf x. It is adjacent to exactly one edge e of F'_t. Why is e in F'? Because it is used to connect some required pair u_i, v_i. So u_i must be in the connected component that got contracted into leaf x, and v_i must be outside. But at time t, u_i was not yet connected to v_i, so the component associated to x must be active at that time, qed.

Now, since F'_t is a tree, the sum over every node $u \in F'_t$ of $\text{degree}_{F'}(u)$ is $2|F'_t| - 2$. Partition the nodes of F'_t into A (active) and B (inactive).

$$\sum_A \text{degree}_{F'}(u) + \sum_B \text{degree}_{F'}(u) \leq 2|A| + 2|B|.$$

But inactive nodes are not leaves, so they have degree at least 2, so $\sum_B \text{degree}_{F'}(u) \geq 2|B|$, and so $\sum_A \text{degree}_{F'}(u) \geq 2|A|$, as desired.