

CS6100: Topics in Design and Analysis of Algorithms

End-semester Examination

Arjun Bharat, CS17B006

17th May, 2021

1. (a) Denote PT as the Steiner tree estimated by the partitioning algorithm. We perform the following algorithm:
 - i. Define $F_i = F \cap R_i, G_i = G \cap R_i, \forall 1 \leq i \leq m^2$, where R_i is one of the equal squares formed by partitioning $[0, 1]^2$ into m^2 congruent squares each of area $\frac{1}{m^2}$. This creates m rows and m columns. We let $m^2 = \frac{n}{\log n}$, so that in a uniformly random case, each sub-region has an average of $O(\log n)$ points within it.
 - ii. Obtain an estimate ST_i as the Steiner tree for the problem $ST(F_i, G_i, R_i)$.
 - iii. Given the sub-Steiner trees for each region, we now have a forest of m^2 individual trees. We need to add the $m^2 - 1$ best edges across these forests by linking them in some fashion, such that we get an overall Steiner tree from these. This would ensure that every vertex of G is connected in the resulting tree, and therefore, we would have a valid Steiner tree consisting of only the vertices in F and G that includes every vertex in G . Note that there are $2^{\binom{m^2}{m^2-1}}$ possible linkings overall.
 - iv. While linking two regions, we simply add an edge from one vertex in one sub-Steiner tree to another vertex in another sub-Steiner tree.
 - v. Since we only want an estimate of the cost, we can take any linking. We add the linking edges in a snake-like pattern, that is, starting from the top left corner, going down to the next region while adding a linking edge, then right and back up till the top row, then right and then down the third column and so on. In general, we go down the odd columns and up the even columns, if columns are numbered 1 to m from left to right. Each linking edge has length $O(\frac{1}{m})$ (worst case occurs when we add a diagonal across a $\frac{1}{m} \times \frac{2}{m}$ rectangle). The total weight of the linking edges is $O(\frac{1}{m}) \cdot O(m^2) = O(m)$.

vi. We estimate

$$PT = \left(\sum_{i=1}^{m^2} \text{cost}(\text{ST}_i) \right) + O(m)$$

- (b) For a uniformly random distribution $X = \{X_1, X_2, \dots, X_m\}$, we see that we have $|F_i|$ to be of the form $\mathcal{B}(n, \frac{1}{m^2})$, which is a binomial distribution. We assume that computing each Steiner Tree takes time $f(\text{ST}_i)$ for a function f . Now, we see that the time taken to link and generate the overall Steiner tree is $O(m^2)$, as adding each edge is $O(1)$. Using the stated result from Yukich's book (stated just after theorem 5.8):

If $f(x) = Ax^B 2^x$, $m^2 = \frac{n}{\sigma(n)}$, $X_1, X_2, \dots, X_{\frac{n}{\sigma(n)}}$ be iid with each being

$\mathcal{B}(n, \frac{1}{m^2})$, and $S = \sum_{i=1}^{\frac{n}{\sigma(n)}}$, then we have $\mathbb{E}[S] \leq 4An\sigma(n)^{B-1} \exp(\sigma(n))$.

(This is proved in Karp and Steele).

Here we have some constant A (because $\text{ST}(F, G) \in O(2^{|F|})$), $B = 0$ for the cost function f , and $\sigma(n) = \log n$. Hence, we have,

$$\mathbb{E}\left[\sum_{i=1}^{m^2} \text{cost}(\text{ST}_i)\right] \leq \frac{4An^2}{\log n}.$$

Hence, we have Expected running time $= O(m^2) + \mathbb{E}\left[\sum_{i=1}^{m^2}\right] + O(m^2) \leq$

$$\frac{4An^2}{\log n} + O\left(\frac{n}{\log n}\right) \implies \text{Expected running time} \in O(n^2).$$

We assume that the Steiner Tree functional obeys the growth bound, and hence, we have $\mathbb{E}[\text{ST}(F, G)] \leq c_1 \cdot |F|^{\frac{1}{2}}$, for some constant c_1 .

By the linearity of expectation, we have, $\mathbb{E}[\text{PT}(X)] \leq \sum_{i=1}^{m^2} \mathbb{E}[\text{ST}_i] +$

$$O(m) \leq c_1 \cdot \sum_{i=1}^{m^2} \mathbb{E}[|F_i|^{\frac{1}{2}}] + O(m) \leq c_1 \cdot \left(\sum_{i=1}^{m^2} \mathbb{E}[|F_i|]^{\frac{1}{2}} \right) + O(m) = c_1 \cdot m \cdot$$

$\sqrt{n} + O(m) = \frac{c_1 n}{\sqrt{\log n}} + \frac{\sqrt{n}}{\sqrt{\log n}} \in O(n)$. (Using Jensen's inequality on the concave function \sqrt{x} .) Now, assuming the closeness to the boundary functional, we would have $\mathbb{E}[\text{ST}(F, G)] \geq c_2 \cdot \sqrt{n} - \mathbb{E}[\text{ST}_B(F, G)] = c_2 \cdot \sqrt{n} - O(1)$, (since $\mathbb{E}[\text{S}_B(F, G)] = O(n^{\frac{d-2}{d}}) = O(1)$). Hence, we have $\frac{\mathbb{E}[\text{PT}(X)]}{\mathbb{E}[\text{ST}(X)]} \leq \frac{c_1 n}{c_2 \sqrt{n} - O(1)}$

- (c) If we consider a perturbed distribution on the parameter, we can argue that each F_i is now a binomial random variable with probability of success at most $\frac{\phi}{m^2}$, using a simple union bound. Therefore, we argue that a scaling factor of ϕ is applied to the success probability, and hence, it suffices to replace $\sigma(n)$ by $\phi\sigma(n)$ the right hand side of the running time bound of $\mathbb{E}\left[\sum_{i=1}^{m^2} \text{cost}(\text{ST}_i)\right]$. Hence, we would have, expected running time bound $= O(m^2) + \frac{4An^{(1+\phi)}}{\phi \log n} \in O\left(\frac{n^{1+\phi}}{\phi}\right)$

2.