Approximation Algorithm

Prof. Palash Dey

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

Week – 08

Lecture 40

Lecture 40 : Randomized Rounding for Uncapacitated Facility Location

Welcome. So, in the last class we have seen we have finished the randomized rounding based algorithm for price collecting standard tree problem and the approximation ratio was $\frac{1}{1-e^{\frac{-1}{2}}}$. And, we have also derandomized that algorithm which is very natural. In this class we will we will see the randomized rounding based algorithm for uncapacitated facility location problem for which we have already seen deterministic rounding based 4 factor approximation algorithm. We will improve the approximation factor for that algorithm for that problem from 4 to 3. So, let us start.

So, it is uncapacitated facility location problem. So, let us briefly recall what was our linear programming relaxation and the dual and how we got a four factor approximation algorithm in the deterministic rounding. So, the primal LP was minimize sum of facility opening costs $f_i y_i$, $i \in F$ plus sum of assignment costs $i \in F$ $j \in D$ set of clients $c_{ij} x_{ij}$ subject to all clients is assigned to some facilities equal to 1 and a client can be assigned to a facility only if the facility is open for all $i \in F$ for all $j \in D$. x_{ij} it is it is 1 only if y_i is 1.

So, this is less than equal to y_i in particular if y_i is 0 x_{ij} has to be 0. And of course, all the variables are non negative $i \in F$ $j \in D$ x_{ij} is greater than equal to 0 for all $i \in F$ $y_i \geq 0$. So, this is a primal LP the dual LP was maximize $\sum v_j$ subject to $\sum w_{ij} \leq f_i$ for all $i \in F$. That means, for every facility the facility opening cost is paid by the distributed among the clients which receive it w_{ij}s are the variables v_js are the variables. and $v_j - w_{ij} \leq c_{ij}$ for all $i \in F$ and $j \in D$ and of course, $w_{ij} \geq 0$ for all $i \in F$ $j \in D$.

F_i's could be negative because F_i's are variables corresponding to this equality constraint. And then as usual we any rounding based algorithms we solved, but in this case we solve both the primal LP and dual LP. So, let (x^*, y^*) and (v^*, w^*) be primal and dual optimal solutions. And let us recall we define the concept of neighborhood. So, we say that for any client j the neighboring facilities $N(j)$ are all those facilities such that $x_{ij}^* \geq 0$.

ok. And again for every $j \in D$ we defined $N^2(j)$ is the set of all clients $k \in D$ such that there exist a facility i in the neighborhood of j such that facility i belongs to the neighborhood of k also ok. How does the deterministic rounding algorithm worked? So, let us briefly recall let us write the pseudo code. So, first we solved and solve both the LPs and got the optimal solutions. Initially all the clients are unassigned and we are opening facilities iteratively to 0 which is a counter maintaining the number of facilities we have opened so far.

So, while C not equal to empty set. That means, we have some clients to assign, then what do you do? $k = k+1$, then we choose a client see client j_k which has the lowest v^* values that we saw is useful in the proof of the approximation guarantee of our algorithm. So, choose $j_k \in C$ that minimizes v_j^* over all $j \in C$. So, once we choose this client then I look at the neighbouring facilities of j_k and open the facility which has the least facility opening cost. So, choose $i_k \in N(j_k)$ that has least facility opening cost in $N(j_k)$ ok.

And then we assign j_k and all unassigned clients in $N^2(j_k)$ to i_k ok. And we remove this from all the clients that we have assigned in this iteration, we remove this from the set of clients which are unassigned $C \setminus N^2(j_k)$, $N^2(j_k)$ includes j_k also and that is it. So, this is the algorithm and then how did we analyze? We showed that total facility opening cost of the algorithm is less than equal to $\sum f_i y_i^*$ which is less than equal to opt and total assignment cost that we have seen that for each client the assignment cost is at most 3 times v_j^* for client j. So, this is less than equal to $3 \sum v_j^*$ which is again less than equal to 3 times opt these are from here. see the summation register is opt and $\sum f_i y_i^{□}$ plus some other term is also opt and using these two we got a 4 factor approximation algorithm.

So, total cost is at most 4 times But this analysis is slightly loose in the sense that we are we have bound the facility opening cost with $\sum f_i y_i^*$ which is less than equal to opt, but we have better bound on opt. We know that this is that opt is actually $\sum f_i y_i^* + \sum c_{ij} x_{ij}^*$. So, that is the modification that we will do in the randomized rounding algorithm. So, that we can use this stronger bound for opt that we have is that opt equal to $\sum f_i y_i^* + \sum c_{ij} x_{ij}^*$. So, this is equal to opt.

So, what is the what is the what change we do in this in this rounding technique? The first one is instead of choosing in this step instead of choosing an unassigned client which has least v_j^* value over all $j \in C$, we choose least v_j^* plus something called C_j^* value. So, what is C_j^*? So, C_j^* so, for client $j \in D$ define C_j^* to be. it is expected cost if I open a random facility if I open a facility as per the probability distribution x_{ij} it is expected cost assignment cost $\sum_{i \in F} c_{ij} x_{ij}^*$. So, if I open a facility i with probability x_{ij}^* and assign it to j then this is the expected assignment cost this is the C_j^*.

So, I pick a j I pick a unassigned client which has minimum $v_j^* + C_j^*$. We will see that this modification will allow us to include C_j^* value here in this analysis we will see. And the second modification is for this to make sense we should not open a facility in a neighboring facility of j_k which has least facility opening cost which will which we should open the facility each open one facility in the neighborhood of j_k with probability $x_{j_k}^*$. So, that modification we do choose j_k not with least opening cost, choose $i_k \in N(j_k)$ with probability $x_{i_k j_k}^*$ to open. So, these are the two modifications we do and now we will show that these are three factor approximation algorithm ok.

So, the claim or theorem the randomized rounding based algorithm is a three factor approximation algorithm. So, let us see the proof. In iteration k, the expected facility opening cost is. So, you see we open a facility $i_k \in N(j_k)$ with probability $x_{i_k j_k}^*$. So, this is this cost is $\sum_{i \in N(j_k)} f_i x_{i j_k}^*$.

Now, this is less than equal to summation. So, our goal is to bring this term. So, for that we use this inequality that $x_{ij}^* \leq y_i^*$ this is $f_i y_i^*$ $i \in N(j_k)$. Now, again as usual if let $j_1, \ldots, j_t$ be the clients picked. at step let us give some name of the steps this is 1 2 3 4.

So, step 5 of the algorithm then by choice n their neighbor roots are disjoint ok. So, total expected facility opening cost is $\sum_{k=1}^{t} \sum_{i \in N(j_k)} f_i y_i^*$ is at most. Now, because this neighborhoods are disjoint this is less than equal to $\sum f_i y_i^*$. So, the facility opening cost expected facility opening cost that bound is same as the bound for the deterministic rounding algorithm. So, now, we will do the expected assignment cost and then we will see an improvement.

So, again let k be any iteration. and $j = j_k$ be the client chosen in this iteration and i the facility opened, then expected assignment cost of j of assigning j to i is $\sum_{i \in N(j)} c_{ij}^* x_{ij}^*$ which is C_j^*. So, here is $j = j_k$ this is the neighborhood of $N(j)$ neighborhood of j. and some facility i is open in this iteration which is picked with probability x_{ij}^* and all the clients in $N^2(j)$ are assigned to i. So, we have seen that this cost the expected assignment cost of j to i is C_j^*.

Now, I want to analyze if there is other for other clients $l \in N(j)$ what is the expected assignment cost. the expected assignment cost of assigning an unassigned client $l \in N^2(j)$ to i is at most. So, why $l \in N^2(j)$? Because there is a neighbouring facility let us call it h which is in $N(j)$. So, h belongs to $N(l)$. So, c_{il} is less than equal to this h plus this h plus this h.

So, this is at most $c_{hl}+c_{hj}$ plus. So, these are deterministic term and the and the c_{ij} which is random and. So, it is expectation is $\sum_{i \in N(j)} c_{ij} x_{ij}^*$ ok. Now, this is equal to $c_{hl}+c_{hj}+C_j^*$. Now, c_{hl} because h is a neighbour of l.

So, $c_{hl} \leq v_l$ and because h is a neighborhood of j also $c_{hj} \leq v_j^*+C_j^*$ this is since h is a neighbor of l and h is a neighbor of j. Now, this is then now $v_j^*+C_j^* \leq v_l^*+C_l^*$ because in the beginning of k-th iteration both j and l was unassigned and we choose the client j which has the minimum $v_j^*+C_j^*$. So, in particular $v_j^*+C_j^* \leq v_l^*+C_l^*$. So, this is from choice j is $2v_l^*+C_l^*$. Hence, total cost is less than equal to total facility opening cost is this $\sum f_i y_i^*$ plus for each client the assignment cost is at most $2v_l^*+C_l^*$.

So, for every client $j \in D$ this is $2v_j^*+C_j^*$. So, by rearranging $\sum_{i \in F} f_i y_i^* + \sum_{j \in D} C_j^* + 2\sum_{j \in D} v_j^*$. this is the primal opt and this is dual opt both are same. So, this is equal to 3 LP opt by strong duality which is less than equal to 3 opt. Hence, this is a 3 factor randomized approximation algorithm ok. So, let us stop here. Thank you.