Approximation Algorithm

Prof. Palash Dey

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

Week – 06

Lecture 29

Lecture 29 : A 4 Factor Approximation Algorithm for Uncapacitated Facility Location Problem

Welcome. So, in the last class we have seen a deterministic rounding based algorithm for price collecting standard tree. So, today we will start discussing another very important problem which is called uncapacitated facility location problem. It is a well studied problem and we will revisit this problem many times in our course to introduce various algorithm design techniques and frameworks. So, today we will see uncapacitated capacitive facility location problem. we will see a deterministic rounding based algorithm which achieves a 4 factor approximation algorithm for this problem.

So, what is the problem input? A set F of facilities and a set D of demands or clients each facility $i \in F$ has facility opening cost f_i and there is an assignment or service cost for each facility in F and demand $j \in D$, there is an assignment cost or serviced cost c_{ij} all are non-negative ok. It is the cost if a facility i needs to serve the demand d or client d it depends on distance and so on. So, what is the goal? The goal is minimize if you decide a set $F^{'}$ of facilities to open then the total cost that you incur is total facility cost.

$f_i, i \in F^{'}$ and once you decide which facilities to open if you want to we want to minimize the assignment cost. So, for that for every demand or client we simply assign it to its nearby or nearest facility. That means, whichever facility has minimum cost of assigning it to the client. So, for every client $j \in D$ minimize $i \in F^{'}$ c_{ij} ok. So, this quantity we want to minimize.

Now in its full generality, so why it is called an uncapacitated facility location problem? Because here the facilities does not have any upper bound on the number of clients that it can serve that is why it is called uncapacitated facility location problem. In the capacitated facility location problem each facility has a upper bound on how many clients it can serve. So, in its full generality this problem is as hard as set cover problem. And so, we will see constant factor approximation algorithm for this problem assuming that the facilities and clients are points in a matrix space. So, we assume facilities and clients are points in a common matrix space .

what does this give us? This gives us triangle inequality. Hence, assignment costs triangle inequality. What do I mean by that? If I have two facilities i, i' their facilities and I have two clients $j, j' \in D$. So, the cost of assigning i to j that is c_{ij} is less than equal to if I assign the cost of assigning i' to j $c_{i'j}$ plus cost of assigning i' and $j'$and cost of assigning i to j' ok.

So, this is what do why what we mean by saying that assignment cost respect triangle inequality ok. So, as usual towards writing a integer linear programming formulation let us have variables. So, we have a variable y_i for every facility $i \in F$ which will take value 1 if we open i. otherwise it takes value 0 and for every assignment we have a variable we have a variable x_{ij} for every facility $i \in F$ and client $j \in D$ which we will take value 1 if we assign client j to facility i. ok otherwise it takes value 0.

So, with this semantics of variables a we it is easy to write down the objective function which is we want to minimize summation of facility opening costs in F $f_i y_i$ plus summation of assignment costs in F x_{ij} and now we need to write down constraints. So, that these variables are forced to take their intended values. So, we write constraints. First every client needs to be assigned to some facility that is exactly one facility and that must be open . So, every client $j \in D$ must be assigned to exactly one facility that is the constraint is for all client $j \in D$ $\sum_{i \in F} x_{ij}$ this should be equal to 1 ok.

And then the next constraint is a facility a client j can be assigned to facility i only if it is open this is the first constraint. The second constraint is a client $j \in D$can be assigned to facility i only if i is open that is for all $j \in D$ for all $i \in F$ x_{ij} if it takes value 1 then y_i must also take value 1. So, x_{ij} is less than equal to y_i. if $y_i$takes value 0 then x_{ij} must take value 0. So, what are the ILP then minimize $\sum_{e \in E} c_e x_e$ plus Minimize total facility opening cost $\sum_{i \in F} f_i y_i$ plus sum of assignment costs $i \in F$ $j \in D$ c_{ij}.

x_{ij} subject to what are the constraints? That for all client $j \in D$ $\sum_{i \in F} x_{ij} = 1$ and for all $i \in F$ $j \in D$ $x_{ij} \leq f_i$ if f_i is 0 x_{ij} must be 0 and for all $i \in F$ and for all $j \in D$ what we have is f_i takes value in 0 1 and x_{ij}takes value in 0 1. So, this is the integer linear programming formulation which is exactly which exactly capture the problem and hence ILP of equal to opt. Next as usual we relax the integrality constraints so that we can solve it. So, we replace it with f_i Now, because this is a minimization problem and f_i's and f_i's appear positively it is enough if I write $f_i$greater than equal to 0. If some f_i takes value more than 1 then I can make that f_i take value 1 and then all the constraints will remain satisfied and if x_{ij}'s I can write x_{ij} is greater than equal to 0.

If I can remove the constraint that x_{ij} cannot take greater than 1 value. If some x_{ij} takes value greater than equal to 1, I can reduce it to take the value 1 that only reduces the objective function and if x_{ij} take value greater than 1 that is that violates this constraint. So, x_{ij} cannot take value greater than 1 first of all and if x_{ij} cannot take value greater than 1 then if f_i some $f_i$takes value greater than 1 I can change that value to 1. and that continue to satisfy all the constraints, but reduces the objective function because all f_is if it is positive if f_i's are 0 then it does not affect the value of the optimization function. So, without loss of generality we can drop the constraints that $f_i$is less than equal to 1 and x_{ij}s less than equal to 1 that is a quite a handful number of constraints LP, because in this problem we will be needing duals.

We will write the dual and we will solve both this primal LP and dual LP, we will take both the optimal solutions to guide us to compute an approximately optimal integral solution. and for writing dual the number of constraints here will correspond to the number of dual variables. So, whenever we try to write dual it may be it may make some sense it may make sense to reduce the number of constraints as far as possible. So, what is the dual of this? Of course, you can write it write it mechanically and that is enough, but let me motivate the dual problem with some assigning some meanings to the variables like here these variables y_i s and x_{ij} s have some meaning. So, let me motivate the dual linear program if based on some meaning.

So, what we do is we are trying to get a lower bound on opt. We want to obtain a lower bound on opt ok. So, for that for the time being let us ignore the facility opening cost. So, if there is no facility opening cost obviously, the best solution is to open all facilities and assign each client to its nearby facility. So, let $v_j, j \in D$ be the cost incurred by client G ok.

So, we are asking that assignment cost let the client pay and that is v_j. So, what will be v_j then? v_j then will be minimum of $i \in F\, c_{ij}$. ok and this is the minimum thing that we that each a client needs to pay. Hence summation $v_j, j \in D$ here $j \in D$ is a lower bound on opt. So, value of opt should be greater than equal to summation v_j.

Now, let us net let us bring this important fact that you know the facility opening cost is non-zero there are facility opening cost. So, the idea is pass distribute the facility opening costs among the clients connected to it. ok that is f i equal to $\sum_{j\in D} w_{ij}$. So, w_{ij} is the is the cost that j pays for opening facility i, j will pay only if j is connected to i. ok and we want w_{ij}s to be greater than equal to 0 ok.

So, client j w_{ij} if j is connected to i. But, then how do we once we have these w_{ij}'s how do we get this w_{ij}'s? The idea is we will write an optimization function to get this w_{ij}'s which satisfy all these constraints. So, w_{ij}'s will be the variables exactly the dual variables and those w_{ij}'s correspond to the dual variables correspond to these constraints. So, here will be w_{ij}'sand here will be v_j's. v_j's will be the will be the variables corresponding to these constraints and then what I want again.

So, from each a client's point of view he needs to pay v_j what is it v_j is $min_{i\in F} c_{ij}+w_{ij}$. Now, how do I ensure that v_j is set to minimum? One way to ensure is to ensure the above equality we have constraint. have a constraint for all $i\in F$ we write v_j is minimum. So, this is less than equal $c_{ij}+w_{ij}$ and we maximize some v_j So, if I want to take the maximum value that v_j can take subject to the constraints that v_j is less than equal to $c_{ij}+w_{ij}$ s, then I need to set v_j to be $min_{i\in F} c_{ij}+w_{ij}$ that is exactly what I want.

So, what are the optimization function then optimization problem maximize $\sum_{j\in D} v_j$ subject to $\sum_{j\in D} w_{ij}$ this is less than equal to f_i for all $i\in F$. and v_j is less than equal to $c_{ij}+w_{ij}$. So, v_j is a variable. So, if we write both on the left hand side then we have got $v_j - w_{ij} \leq c_{ij}$ this is for all $i\in F$ for all$j\in D$. ok and w_{ij}'s are greater than equal to 0 that is all.

So, this is the dual LP and this is the primal LP. So, if you mechanically write the dual of this primal LP you will get that LP and an important point to note here is that the dual variables corresponding to this equality v_j's because it is equality $v_j$can take both positive and negative sign. So, on the other hand because for w_{ij} this constraints are inequality constraints. So, w_{ij} should be greater than equal to 0. This is what we can see here also that w_{ij} are greater than equal to 0 for all $i\in F$ for all $j\in D$, but we do not have any restriction on the values of $V[G]$'s that they can take.

So, this is a dual linear program and in the next lecture we will see how by taking a primal optimal solution and a dual optimal solution we use them as a guide to design a four factor approximate solution for this problem ok. So, let us stop here. Thank you.