Approximation Algorithm

Prof. Palash Dey

Department of Computer Science and Engineering

Indian Institute of Technology, Kharagpur

Week – 01

Lecture 04

Lecture 04 : Dual Rounding: An Approximation Algorithm for Weighted Set Cover

Welcome. So, in the last class we have seen the dual linear program for the linear programming relaxation of set cover and we have seen two crucial few crucial results for linear programming duality in the context of set cover. One is weak duality theorem, it says that any dual solution of the dual linear program provides a lower bound on the primal objective. Second is strong duality, it says that if both primal and dual linear programs are feasible, then their optimum values coincide The third one is complementary slackness. It says that if I have a pair of solutions primal and dual and for every primal solution which is non-zero the corresponding dual constraint is tight and also if for every dual variable which is non-zero the corresponding primal is tight. if both holds all these conditions hold then x star and y star the solutions that we started with is an optimal solution and it is a characterization of the optimal solution.

And with this let us continue our rounding for dual linear program for set cover, dual for set cover. I repeat we are only giving a very high level overview of linear programming duality. I strongly recommend that you study linear programming and duality very well from any standard material and there are lot of material available on online. This is this high level overview is not at all a substitute of in-depth study which we will require throughout this course of approximation algorithm.

So, let us recall what is the dual linear program, we are charging each element e_i some amount y_i and we want to maximize that charge. So, maximize $\sum_{i=1}^{n} y_i$ ok and we cannot charge any element any set of elements whose total sum of charges is more than the weight of the set. So, subject to for all set $j \in [m]$ if I sum the weight charges equal to $\sum_{i \in [n] e_i \in S_j} y_i \le w_j$ and for all $i \in [n], y_i \ge 0$. So, this is the dual linear program now we will round it to get an approximation algorithm for set curve. So, what is the algorithm? The algorithm is very simple.

So, we solve it and find an optimal solution recall linear programs can be solved in polynomial time that is a breakthrough result in theoretical computer science. the popular simplex method although it is practically very fast, but it is not a polynomial time algorithm. So, let $(y_i^*)_{i\in[n]}$ be any optimal solution of the above dual linear program. And now we want to pick sets which are which is a solution and sets correspond to constraints. So, we pick those sets for which the corresponding constraint is tight.

Again what is tight? Tight means that this inequality holds with equality. So, let I be those sets S_j such that this constraint corresponding constraint is such tight such that $\sum_{i\in[n]:e_i\in S_j} y_i^* = w_j$. So, let I be the set of such sets and our algorithm outputs I. So, few things we need to prove first of all it is not clear that I is indeed a set cover let alone it is a approximately optimal set cover. So, let us first prove that I is a set cover.

I is a set cover proof, it is a proof by contradiction. So, suppose I is not a set cover. that means, there exist an element which is not covered by the sets in I. that means, that for all sets which contain it is corresponding inequality is not tight. So, we need to assume that for every element there exist a set which contains it otherwise it is a clearly a no instance it is a infeasible solution.

So, I is a set covered assuming the instance is feasible that is every element of the universe U belongs to some sets. This requirement you can see in the linear program also. If there exist an element which does not belong to any set, then that element we can set it to as high up value as possible which drives the maximum value of this linear program as high as possible and hence this linear program will be unbounded. So, there will be no optimal. and this is this is again a caviar this again follows from linear programming duality.

If the primal linear program is infeasible then the dual linear program must be unbounded and it has to be because if it is not unbounded again it follows from weak duality. So, now, let us come back to the proof. Suppose e_i is an element which is not covered by I, then this implies that for all $j\in[m]$ for all sets S_j such that $e_i\in S_j$ ok. For all such sets $e_i\in S_j$ all such sets which contains e_i the inequality is not tight inequality is strict $\sum_{i\in[n]:e_i\in S_j} y_i^* < w_j$. Now, you see that because these take real numbers and because all inequalities where y_i^* appear is strict that means, these are less than not equal to that means, there exist some positive real number epsilon with which we can increase the value of y_i^* and still all inequalities hold.

This implies that there exists an epsilon greater than such that we can increase the value of y_i^* by ϵ without violating any constraint. However, if so we got another feasible solution with higher objective function value. So, this contradicts our assumption that y_i^* . So, let us use some other notation say $(y_i^*)_{i\in[n]}$. is an optimal solution optimal dual solution ok.

because it is a maximization problem and if y_i^* is an optimal solution no solution can have higher value, but we have obtained another solution whose sum $\sum_{i=1}^{n} y_{i'}$ ϵ more than the value at y_i^* . So, how come y_i^* be an optimal solution. So, this shows that we cannot have any element e_i which is uncovered, hence I must be a set cover, hence I must be a set cover. So, our algorithm outputs a valid set cover. Next we need to show that it is an approximately optimal solution.

So, we claim next claim our algorithm achieves an approximation factor of f, where f is the maximum frequency of any element of the universe. we call what is the frequency of an element? It is the number of sets where that element appears proof ok. So, what is ALG? What is the cost of my solution? ALG is summation of the sets i is the set that is output by my algorithm. So, $ALG=\sum_{j\in[m]:S_j\in I} w_j$ ok.

is this. So, this is $\sum_{j\in[m]:S_j\in I} w_j$ now instead of w_j we write this w_j is greater than equal to this. and see this is equal to because we are we have picked sets which are tight. So, for every set $S_j\in I, w_j=\sum_{i\in[n]:e_i\in S_j} y_i^*$ ok. So, for every set $S_j\in I$ I am summing over the y_i value of the elements in S_j . So, again we have a double sum and as usual we will try to rewrite the double sum exchanging the double sum and we will get useful results.

So, you want to write from i equal to 1 to n. Now you see look at from every elements perspective how many times their y_i^* appears for an element e_i . It is appearing for those many sets which pick which contain e_i and part of the solution. So, this is you see this is the cardinality of those sets $j\in[m]$ such that $S_j\in I$ and $e_i\in S_j$. So, this many times it is appearing y_i^* ok.

Now, this number is the number of times e_i appear in the set in the solution set. So, this number is less than equal to f which is the maximum frequency of any element. So, this is less than equal to $\sum_{i=1}^{n} f\, y_i^*$ take f out. $f\sum_{i=1}^{n} y_i^*$ this is dual opt and dual opt is equal to primal opt by strong duality which is less than equal to opt. So, we have shown that ALG is less than equal to f times OPT.

Hence our algorithm has an approximation factor of f. Now, can we improve this analysis? It turns out not and we have we show some interesting result from complementary slackness. Let I be the solution which is a collection of sets let I be the solution of the dual rounding algorithm and I' the solution of the deterministic primal rounding algorithm then $I' \subseteq I$. That means, the dual rounding algorithm picks all sets that the primal rounding algorithm picks and maybe more proof. Let us recall what was the primal rounding algorithm.

So, i prime. So, let $(x_j^*)_{j\in[m]}$ and $(y_i^*)_{i\in[n]}$ be primal and dual ah optimal solutions ok. And let us recall what was I' and what was what is I. i is those sets S_j such that the dual constraint is tight means $\sum_{i\in[n]:e_i\in S_i} y_i^* = w_j$. this is I and what is I' ? I' is all those sets S_j such that the optimal values x_j^* this is greater than equal to $\frac{1}{f}$ ok.

So, that is the thing. Now, recall complementary slackness says that because both are optimal solutions whenever primal is non-zero that means, dual is dual constraint is tight whenever the primal some primal variable is non-zero the corresponding dual constraint is tight. Recall for every primal variable we have a corresponding dual constraint and for every dual variable we have a corresponding primal constraint. complementary slackness conditions we have for all $j\in[m], x_j^* \geq \frac{1}{f}$ this implies that the corresponding constraint must be tight. The corresponding dual constraint must be tight $\sum_{i\in[n]:e_i\in S_i} y_i^* = w_j$.

This implies that for all $j\in[m], S_j \in I'$ implies $S_j \in I$. which is same as saying I prime is a subset of I ok. So, the solution the total weight of the sets of the dual linear programming relaxation algorithm cannot be less than the primal deterministic rounding of the primal linear programming relaxation. and this follows from complementary slackness conditions. So, in the next lecture we will continue and see some more use of complementary slackness and primal dual linear programs for designing approximation algorithms. Thank you.