Mid-semester Examination: CS60023: Approximation and online algorithms

LTP 3-0-0: Credits 3: Time 2 hours: Spring 2019
You must answer all questions totalling 100 marks.
All proofs and arguments must be complete and clear.
Follow lucid writing style, using suitable notation and maintaining rigour.

February 20, 2019, 9 am - 11 am

(1) Show that the minimum cost of a perfect matching on the set V' of odd degree vertices in an MST of $G(V, E)$ is at most the half of OPT where OPT is the minimum cost of a metric TSP tour of the complete graph $G(V, E)$. Here the positive edge weights in $G(V, E)$ obey the triangle inequality. (5+5+5 marks)

[Solution: See moodle notes or the texts.]

(2) State the K-centre problem precisely and the greedy approximation algorithm discussed in class for the K-centre problem for a complete weighted graph $G(V, E)$, with positive edge weights obeying the triangle inequaliy.

Once this algorithm identifies K centres, suppose one vertex v, has its cluster centre $c(v)$. We know that the distance between v and $c(v)$ is at most $2 \times OPT$. [Here, OPT is the optimal radius of the K-centre problem for the given graph.] Let the nearest neighbour of $c(v)$ in the $G(V, E)$ be w. Estimate an upper bound for the distance between v and w in terms of OPT. (5+5 marks)

[Solution sketch: The two cases must be argued fully; in one case each center computed by the approximation algorithm lies in a distinct cluster of the optimal solution. The other case is the complementary case. For the second part, use triangle inequality.]

(3) Show that we can design a polynomial time f-factor algorithm for the *weighted set covering* problem if each element is covered in at most f sets. To show this you may use the primal rounding technique, that is, when the optimal solution has a value at least $\frac{1}{f}$ for an indicator variable, we choose the corresponding indicated set in the set cover. You must first clearly state and explain the formulations of the primal and dual LPs. Secondly, you must show that the collection of sets picked up is indeed a set cover. Finally, you have to argue why the set cover computed is an f-factor approximation. (6+4+5 marks)

[Solution sketch: Primal indicator variables that are at least $\frac{1}{f}$ are hiked up to 1, at most f-fold. Thus, the approximation factor is f. Feasibility in the primal inequalities (one for each element), where the element is in at most f sets, implies at least one of the f indicator variables in the optimal fractional relaxed primal LP must have value at least $\frac{1}{f}$.

That very set will cover the element.]

(4) For the local search (offline) approximation algorithm for scheduling n jobs on m identical machines, where p_i is the time required for the job i, $1 \leq i \leq n$, we start with some schedule for all the n jobs and then undertake *local actions* in several steps to alter the schedule till the termination of the algorithm. The local action is that of transfering the latest ending job (say b) to the machine that has processed least. How many times can such a local step be repeated until termination? [Note that with each local step of the transfer of a job from one machine to another, the total processing time is improved. Argue whether the same job can be transfered repeatedly from machine to machine.]

When no more transfer of any job is possible, argue that all machines must have been busy at the starting time of the last ending job.

Develop lower bounds for the total optimal time OPT for processing the jobs.

Estimate an upper bound on the total time required to complete all jobs in terms of OPT. (3+3+4+5 marks)

[Solution sketch: Natural lower bounds for the total completion required time are (i) the duration of the longest job, and (ii) the average duration of all jobs. For job b that ends last when we cannot move any jobs and the algorithms terminates, we can say at the starting time $S_b = C_b - p_b$ of b, all other machines must have been busy. From time 0 through S_b, all machines must have been busy, so S_b is at most the average, which is at most OPT. Also, completion at C_b after instant S_b, of duration p_b, is at most OPT. So, the total duration is at most $2 \times OPT$.]

(5) Construct a "tight" example for the *multiway cut* approximation algorithm where an approximation ratio approaching $2 - \frac{2}{k}$ is attained. In this problem, a set of k vertices are specified in a weighted undirected graph, and we wish to minimize the weight of a multiway cut that separates the k specified vertices. (10 marks)

[Solution: See Vazirani's text.]

(6) In the maximum coverage problem, we pick k sets from a collection of l given subsets $S_1, S_2, S_3, ..., S_l$, of a universal set U, where $|U| = n$, so that the k sets picked cover the maximum number of elements from U.

Consider the greedy method where every next set picked by the algorithm covers the maximum possible number of yet uncovered elements from U. If OPT is the maximum coverage possible by some k sets of the $l \geq k$ given subsets of U, then show that the greedy method covers at least $(1 - \frac{1}{e})OPT$ elements. (6+9 marks)

[See uploaded notes in moodle.]

(7) The (unweighted) vertex covering problem may be viewed as a set covering problem where we need to cover all the edges incident at each of the vertices. So, for each vertex $v \in V$ of the given unweighted undirected graph $G(V, E)$, we need to form its set $E(v)$ of its incident edges from the set E. Show that using the greedy cardinality set cover heuristic we can compute a vertex cover for $G(V, E)$ which has at most $O(\log n) \times OPT$ vertices, where OPT is the size of the minimum cardinality vertex cover for $G(V, E)$. (10 marks)

[This is also an exercise in Vazirani's text.]

(8) State the primal minimization LP in the form where we minimize $c^t x$ given $Ax \geq b$ for the n-dimensional vector x with non-negative components, with m inequalities on the n variables in x are represented by an $m \times n$ matrix A, and an m-dimensional vector b. Develop its dual maximization LP with m variables in the dual vector y so that the objective functions of the primal-dual pair obey the weak duality and strong duality conditions.

For the facility location problem too, state its primal relaxed minimization LP based on the Integer Linear Program (ILP) for minimization of the sum of opened facility costs as well as costs of connections for all clients to the opened facilities. Develop a dual maximization LP for the primal relaxed minimization LP.

(4+6 marks)

[See notes and the text books.]