Approximation algorithms for discrete stochastic optimization problems

David B. Shmoys

Exercises

1. Consider the 2-stage stochastic set covering problem in the polynomial-scenario model. Show that this problem can be reformulated as (or equivalently, reduced to) the deterministic set covering problem (albeit with a larger family of sets over a larger ground set).

2. In the deterministic minimum multicut problem, the input consists of a graph where each edge has a specified non-negative cost, and pairs of nodes $(s_1, t_1), \ldots, (s_k, t_k)$; the aim is to select a subset of edges of minimum total cost so that after deleting those edges, none of the given pairs has both nodes in the same connected component. For the special case in which the input graph is a tree, there an LP-based algorithm (using the most natural LP) that has a performance guarantee of 2.

   (a) Formulate a 2-stage stochastic optimization problem with recourse that is a natural extension of this problem.

   (b) Derive as good an approximation algorithm for this problem as you can.

3. Consider the 2-stage stochastic uncapacitated facility location problem, and derive a $2\rho_{\mathsf{UFL}}$-approximation algorithm in the polynomial-scenario model, where ρ_{UFL} is the performance guarantee for the deterministic varaint. (Here are some hints to get you started. This will be a variant on the LP rounding technique, though a bit more complicated. Take the optimal solution $x^*_{A,ij}$ and decompose it into $x^I_{A,ij} + x^{II}_{A,ij}$, where $x^I_{A,ij} \leq y^*_i$ and $x^{II}_{A,ij} \leq y^*_{A,i}$. Use this to conclude, for each scenario A and client j that either

$$\sum_i x^I_{A,ij} \geq 1/2,$$

   or else

$$\sum_i x^{II}_{A,ij} \geq 1/2.$$

   For each client j, let S_j be those scenarios for which the above inequality holds for the first stage. This will allow us to devise a stage I input, in which we must simulate the fact that S_j corresponds (roughly) to the demand that must be satisfied in stage I, and yet must also be able to get a feasible fractional solution for this deterministic LP relaxation. From the fractional solution, we can obtain an integer one for the (deterministic) stage I problem. Repeat for stage II. Piece both parts together.) Where do you use that it is the polynomial scenario model?

4. The boosted sampling technique can be generalized to apply to a broad range of optimization problems. Consider the following general optimization problem; there is a universe U of requirement, and there is a set X of elements that can be purchased. For any $F \subseteq X$, let $c(F)$ denote the (linear) cost of F. For any subset $S \subseteq U$, the possible feasible solutions is a set $Sol(S)$ which is a subset of the power set of X. In the deterministic setting, the aim is, for a specified set S, to choose A member of $Sol(S)$ of minimum cost; let $OPT(S)$ denote an optimal solution for S.

   The optimization problem must be *subadditive,* in the following sense. If S and S' are two requirement sets, and F and F' are feasible solutions for them, respectively, then $S \cup S'$ defines another set of requirements, for which $F \cup F'$ is a feasible solution.

   The 2-stage variant again has a probability distribution over subsets of requirements, and allows one to buy some elements $e \in X$ in stage I based only on the distributional information, at a cost of c_e, and then after there is a realized set of requirements drawn from the distribution, one can purchase additional elements $e \in X$ at a cost of λc_e.

The next key element is a cost-sharing mechanism that is given with respect to an approximation algorithm for the deterministic problem. We say that an α-approximation algorithm $\mathcal{A}$ for the optimization problem admits a *β-strict cost sharing mechanism* if there is a function $\xi : 2^U \times U \mapsto \mathbb{R}_{\geq 0}$ such that for every $S, T \subseteq U$ with $S \cap T = \emptyset$, (i) $\xi(S,u) = 0$ for $u \notin S$; (ii) $\sum_{u \in S} \xi(S,u) \leq c(OPT(S))$; and (iii) there is a procedure $\mathsf{Aug}_{\mathcal{A}}$ that augments the solution $\mathcal{A}(S)$ constructed by $\mathcal{A}$ on input S to a solution in $Sol(S \cup T)$ incurring cost $c\big(\mathsf{Aug}_{\mathcal{A}}(S,T)\big) \leq \beta \sum_{u \in T} \xi(S \cup T, u)$.

(a) Show that the minimum-cost rooted Steiner tree problem falls into this framework in which we can provide the requisite algorithms with $\alpha = \beta = 2$.

(b) Provide a boosted sampling algorithm for this general framework that yields an $(\alpha+\beta)$-approximation algorithm for the 2-stage stochastic optimization problem in the black-box model.

5. In the vertex cover problem, the input is a graph $G = (V, E)$ in which each vertex v has a nonnegative weight w_v; the aim is to select a minimum-cost subset of vertices such that for each edge at least one of its endpoints has been selected. There is a well-known primal-dual approximation algorithm for this problem. (A primal-dual algorithm uses the LP framework for its analysis, but does not require solving the LP; for a minimization problem, if we find an feasible integer solution along with a feasible dual solution for its LP relaxation, and the cost of integer solution is at most α times the (dual) cost of the feasible dual solution, then one can conclude that the resulting algorithm has a performance guarantee of α.) In the dual problem, one has a value y_e for each edge, and a solution is feasible if, for each vertex, the sum of the values associated with its incident edges is at most w_v; the objective is to maximize the sum of the dual variables. The primal-dual algorithm works as follows, always maintaining a feasible dual solution: start with all dual variables $y = 0$; choose any edge e, and increase its dual variable until the constraint for one of its endpoints becomes tight; add each vertex with a tight dual constraint to the primal solution (i.e., include it in the cover), delete that vertex (or vertices), and all incident edges, and repeat until no edges remain.

   (a) Prove that this is a 2-approximation algorithm for the deterministic vertex cover problem.

   (b) Generalize this algorithm to derive a 2-approximation algorithm for 2-stage stochastic vertex cover problem in the polynomial-scenario model.

6. Consider the following alternative 2-stage stochastic generalization of the scheduling problem $1|r_j|\sum w_j U_j$ (which might call an augmentation version). More precisely, we are given a set of n jobs, each with a release date r_j before which we cannot process the job, along with a due date d_j by which time we are to complete processing the job; furthermore, each job j has a corresponding weight w_j. In the deterministic variant of the problem, we want to schedule the jobs so as to minimize the total weight of the jobs that do not complete on time, or equivalently, maximize the weight of the jobs that are scheduled within their specified time-window. In this 2-stage stochastic variant, we also have a probability distribution over subsets of jobs that specifies which jobs are active. Suppose that in the first stage, we select a set of jobs that we are committed to serve. In the second stage, for a given scenario, we must schedule each job selected in the first stage, and we may augment this solution by scheduling additional jobs that are active in this scenario. We wish to maximize is the total expected profit (where it is now natural to assume that the profit obtained for an instance in the second stage is less than the corresponding profit in the first).

   Prove that if there is a ρ-approximation algorithm for the augmentation 2-stage stochastic $1|r_j|\sum w_j U_j$, then there is a ρ-approximation algorithm for maximum independent set problem (which is the problem of selected a maximum-size subset of nodes such no pair of them are adjacent). (And since we know strong inapproximability results for the maximum independent set problem, these yield identically strong inapproximability results for this 2-stage stochastic optimization problem.) (Hint: construct a reduction from the maximum independent set problem, and let jobs correspond to nodes, and scenarios correspond to edges.)

7. Complete the details for the result of Charikar, Chekuri, and Pál that shows that the optimal solution for the sample average approximation with polynomial samples for the 2-stage stochastic set covering problem is a near-optimal solution for the true underlying distribution.