Stochastic Combinatorial Optimization: A Survey of Recent Results

Anthony Man-Cho So
Stanford University

The Hong Kong University of Science and Technology
Theory Seminar
22 September 2006

Motivation

- Decision-makers often need to cope with an uncertain environment.
  - News vendor
  - Airline fleet assignment
  - Facility location
  - ...
- A bad decision now may cost a lot later.

Motivation

- To make a “good” decision, one has to:
  - acquire some knowledge of the underlying distribution (modeling)
    - experience
    - market research
  - “optimize” based on this knowledge (optimization)
- In this talk, we will focus on the latter aspect, assuming that we “know” the distribution.

Setting

- So what are we trying to optimize?
  - Let X be the set of possible actions, (Ω, B, P) be the underlying probability space.
  - The outcome is given by $f : X \times \Omega \to R_+$ (e.g. cost).
    - In particular, for each $x \in X$, $f(x,\omega)$ is a random variable.
  - Choose an $x^* \in X$ s.t. $\Phi(f(x^*,\omega))$ is minimized, where Φ is some functional on RVs.
    - expectation
    - max

An Example

	Good ($p=2/3$)	Bad ($p=1/3$)
A	3	3
B	2	4

- It is easy to see that:
  - action B minimizes expected cost; while
  - action A minimizes worst-case cost
- Note that we need to pick one action and stick with it, i.e. we cannot wait-and-see.

Model

- In this talk, we will focus on a class of stochastic programs called "2-stage models with recourse".
  - One commits to some initial action $x \in X$, incurring a cost of $c(x) \geq 0$.
  - Then, the uncertain parameters ω are revealed.
  - One can then respond by taking some further action (recourse) $r \in X'$, thus incurring a cost of $q(x, r, \omega) \geq 0$.

Model

- Our objective is:

$$\min_{x \in X} c(x) + E\left[q'\left(x, \omega\right)\right]$$

where $q'\left(x, \omega\right) = \min_{r \in X'} q\left(x, r, \omega\right)$

- This class captures a wide range of applications.

Example: Stochastic Set Cover

- Universe $U=\{e_1,\ldots,e_n\}$
- Sets $S_1,\ldots,S_m \subseteq U$ with first stage costs $c_{1,i}$ and second stage costs $c_{2,i}$
- Scenario $A \subseteq U$ with prob. p_A; hence $\Omega = 2^U$

Example: Stochastic Set Cover

- The problem:
  - First stage: buy some sets
  - Second stage: after a scenario A is revealed, buy some more sets (if necessary) to ensure that A is covered
  - Goal: minimize total expected cost
- Note that the usual (deterministic) set cover problem requires one to cover U.

Example: Stochastic Set Cover

- Here is the LP relaxation:

$$\begin{aligned} \min \sum_i c_{1,i} x_i + \sum_{A,i} p_A c_{2,i} r_{A,i} \\ \text{s.t.} \quad \sum_{i:e\in S_i} x_i + \sum_{i:e\in S_i} r_{A,i} \geq 1 \quad & \text{for all } A, e \in A \\ x_i, r_{A,i} \geq 0 \quad & \text{for all } A, i \end{aligned}$$

- In general, there are exponentially many variables and constraints!

Example: Stochastic Set Cover

- Equivalently,

$$\min \sum_i c_{1,i} x_i + \sum_A p_A q'(x, A) \quad \text{s.t. } x \geq 0$$

where

$$q'(x, A) = \min \sum_i c_{2,i} r_{A,i}$$

$$\text{s.t. } \sum_{i:e \in S_i} r_{A,i} \geq 1 - \sum_{i:e \in S_i} x_i \quad \text{for all } e \in A$$

$$r_{A,i} \geq 0 \qquad \text{for all } i$$

How to Solve It?

- There have been significant research efforts in the OR community on algorithms for the 2-stage models.
  - However, there are no time bounds for most of the proposed algorithms.
- Only fairly recently, people have started to look for provably fast algorithms for these problems.

Obstacles to Fast Algorithms

- Intuitively, there can be a lot of different realizations, so that even evaluating the objective function at a given $x \in X$ may be a problem!
- Theorem (Dyer, Stougie 2003) 2-stage stochastic programming with discrete distributions on the parameters is #P-hard.

Obstacles to Fast Algorithms

- How about reducing the number of scenarios?
  - Replace the original objective:

$$\min_{x \in X} c(x) + E\left[q'\left(x, \omega\right)\right]$$

by its sample average approximation (SAA):

$$\min_{x \in X} c(x) + \frac{1}{N}\sum_{k=1}^{N} q'\left(x, \omega^k\right)$$

where $\omega^1, \ldots, \omega^N$ are i.i.d. samples from the distribution (assuming we can do that).

Obstacles to Fast Algorithms

- By LLN, the SAA converges to the true objective pointwise.
- Thus, the question is, how big should N be in order to get, say, ε-close (additively) with probability $1-\delta$?
  - For discrete X, $N \approx poly\left(\log|X|, \varepsilon^{-1}, \log\delta^{-1}, V\right)$, where V is the max variance of the RVs $q'(x,\omega)$ over all $x \in X$ (Kleywegt et al. 2001).
  - However, V needs not be polynomially bounded in the input size, and can be hard to compute.

What Can We Do?

- Restrict the problem!
- First Idea: Too many scenarios? Use only succinctly representable distributions.
  - Scenario Model: The set of scenarios and their probabilities are explicitly given.
    - Can take time polynomial in the size of the set.
  - Independent Decisions Model: Each scenario is a set of elements, each of which is chosen independently with some probability.

SSC Under the Scenario Model

- Solve the LP relaxation (polynomial size).
- Now, an element e is covered either:
  - at least $1/2$ by the variables x_i, or
  - at least $1/2$ by the variables $r_{A,i}$ in every A s.t. $e \in A$
- Let

$$E = \left\{ e : \sum_{i:e\in S_i} x_i \geq \frac{1}{2} \right\}$$

Then, $2x$ is a fractional set cover solution for the instance with universe E.

SSC Under the Scenario Model

- Suppose that we have a procedure for set cover that gives $\rho \cdot OPT_{DET}$ wrt the LP relaxation.
- Then, we get an integral cover for E with cost at most $\rho \cdot \sum_i 2c_{1,i}x_i$.
- Similarly for $A \backslash E$ for each scenario A.
- Thus, we get an 2ρ-approximation algorithm.

What Else Can We Do?

- The above scenario models are very restrictive.
- What we really want is the black-box model, where one can sample from the underlying distribution.

What Else Can We Do?

- Second Idea: Dependent on some unnatural parameter V? Replace it with a more natural one.
  - Inflation factor: The second stage cost of a set S is $\lambda_S(\omega)$ times more expensive than its first stage cost. Let $\lambda = \max\{1, \lambda_S(\omega)\}$.
  - Intuitively, λ controls the max variation of the objective.
  - Also, while V can be hard to compute, λ is a parameter of the problem.

Recent Breakthrough I

- Gupta, Pál, Ravi, Sinha: STOC 2004
  - Handles several stochastic network design problems (Steiner tree, facility location, etc.).
  - Fixed inflation factor $\lambda \equiv \lambda_S(\omega)$.
  - Basic framework:
    - Draw λ samples to obtain clients $D_1, \ldots, D_\lambda$.
    - Apply known approximation algorithm on the problem instance $D = \bigcup D_i$ to get a first-stage solution.
    - Once the set of clients is realized, augment.

Recent Breakthrough II

- Shmoys, Swamy: FOCS 2004
  - Works with a broad class of 2-stage stochastic linear programs.
  - Allow scenario-dependent inflation factor.
  - Adaptation of the ellipsoid method.
    - Recall: hyperplane cut is generated by (i) violated inequality, or (ii) subgradient of the objective function at the current ellipsoid center.
    - For (ii), one can use an approximate subgradient instead.

An Illustration (from SS'04)

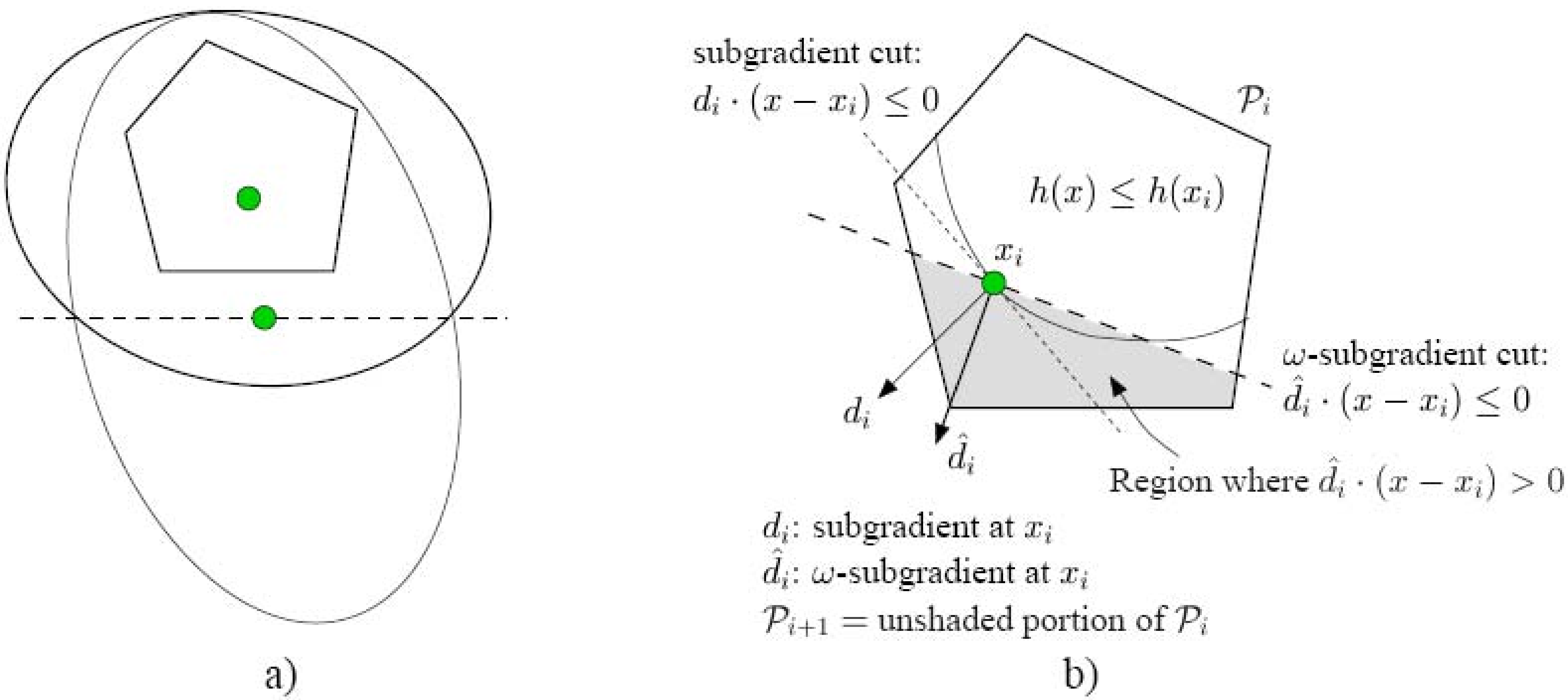


Figure 1: a) Cut derived from a violated inequality. b) Subgradient and ω-subgradient cuts.

Main Ideas of the SS Algorithm

- The approximate subgradient can be computed in randomized poly-time via sampling.
- Convergence can still be guaranteed.
- Finally, use the feasible ellipsoid centers to generate a near-optimal solution with high probability.

Solving SSC w/ the SS Algorithm

- Use the SS algorithm to solve the LP relaxation.
  - Running time: an $(1+\varepsilon)\cdot OPT$ solution with probability at least $1-O(\delta)$ in time
    $$poly\left(input\ size, \lambda, \varepsilon^{-1}, \log\delta^{-1}\right)$$
  - The dependence on λ is necessary in the black-box model.
- Observe that the previous rounding procedure only requires us to examine the first-stage variables!

Recent Breakthrough III

- Charikar, Chekuri, Pál: APPROX 2005
  - Handles general 2-stage stochastic programs.
  - Establish approximation quality for the <u>exact minimizer</u> of the SAA:

$$\min_{x \in X} c(x) + \frac{1}{N} \sum_{k=1}^{N} q'\left(x, \omega^k\right)$$

  under the assumptions that:

  - there exists an $\phi \in X$ s.t $c(\phi) = 0$ and $q'(x, \omega) \leq q'(\phi, \omega)$
  - there exists an $\lambda \geq 1$ s.t. $q'(\phi, \omega) - q'(x, \omega) \leq \lambda c(x)$

Recent Breakthrough III

- An $(1+O(\varepsilon))\cdot OPT$ solution with probability at least $1-O(\delta)$ needs

$$poly\left(\log|X|, \lambda, \varepsilon^{-1}, \log\delta^{-1}\right)$$

  samples.
- Note that this says nothing about how to solve the SAA.

Recent Breakthrough III

- Suppose that we can only solve the SAA approximately, say by a factor of γ.
- We can get an $(1+O(\varepsilon))\cdot\gamma\cdot OPT$ solution with probability at least $1-O(\delta)$ by solving $k=O\left(\varepsilon^{-1}\log\delta^{-1}\right)$ SAAs, each requiring

$$poly\left(\log|X|,\lambda,\varepsilon^{-1},\log\delta^{-1},k\right)$$

samples.

Why Expectation?

- To a risk-averse person, expectation could be a bad idea.
- Recently, Dhamdhere et al. (FOCS 2005) proposed to consider the robust problem:

$$\min_{x \in X} c(x) + \max_{\omega \in \Omega} q'(x, \omega)$$

- They proposed approximation algorithms for several network design problems under the scenario model.

Sampling and Robustness

- Sampling does not work for the min-max problem.
- Can one incorporate some robustness in the black-box model?
  - In some sense, YES! (S., Zhang, Ye: APPROX 2006)

Motivation of a Risk Measure

- How do the aforementioned problems handle robustness?

$q'(x,A_1)$	$q'(x,A_2)$	...	$q'(x,A_m)$
p_1	p_2	...	p_m
1	1	1	1
0	0	0	$1/p_m$

- Assign more weight to a risky (i.e. costly) outcome.

Motivation of a Risk Measure

- For each $x \in X$, consider a weighing function $f_x : \Omega \to R_+$ with $E_P[f_x]=1$.
- Since $E_P[f_x]=1$, we can define a new probability measure $Q_x=f_xP$.
- Consider the problem:

$$\min_{x \in X} c(x) + E_P\left[f_x(\omega) q'(x,\omega)\right]$$
$$= \min_{x \in X} c(x) + E_{Q_x}\left[q'(x,\omega)\right]$$

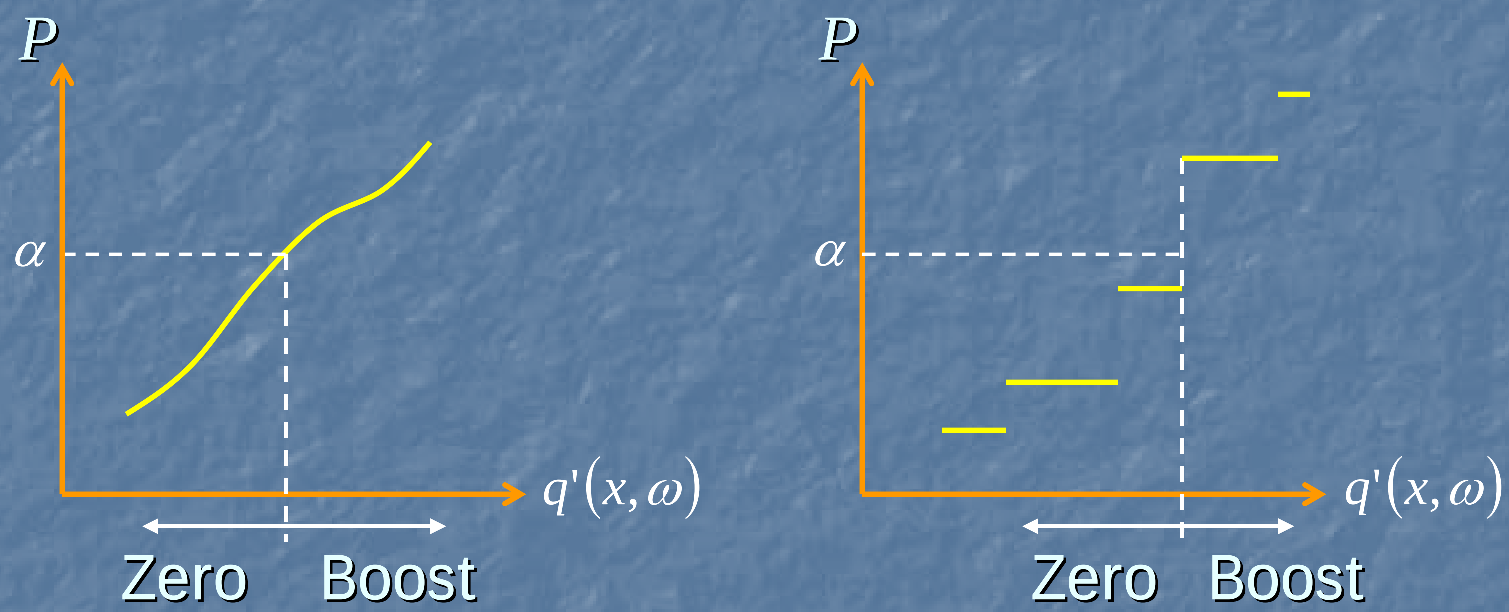
Motivation of a Risk Measure

- What would be a good choice of f_x? It should:
  - make sense
  - computationally tractable
- Let $\alpha \in [0,1)$ be the risk-aversion level. Try:

$$Q = \left\{ f : 0 \le f(\omega) \le \frac{1}{1-\alpha}; E_P[f] = 1 \right\}$$

$$f_x = \arg\max_{f \in Q} E_P[f(\omega) q'(x,\omega)]$$

What Does f_x Do?



Continuous Distribution   Discrete Distribution

$$f_x = \arg\max_{f \in Q} E_P\left[f(\omega) q'(x,\omega) \right]$$

Further Observations

- Observe that when $\alpha = 0$, we have expectation; when $\alpha \uparrow 1$, we have max.
- Note that we can also specify our favorite Q_x without using any weighing function.
  - Under certain conditions, such an Q_x will give rise to a weighing function (in particular, the weighing function is the so-called Radon-Nikodym derivative of Q_x wrt P).

How to Optimize?

- We have a lot of weighing functions – one for each $x \in X$!
- Also, determining the cutoff does not seem obvious if we can only sample from the black-box.

A Representation Theorem

- The following theorem due to Rockafellar and Uryasev (2002) comes to our rescue.

$$E_P\left[f_x(\omega)q'(x,\omega)\right] = \min_{\beta\in R}\left\{\beta + \frac{1}{1-\alpha}E_P\left[\left(q'(x,\omega)-\beta\right)^+\right]\right\}$$

- To simplify notation, write:

$$\varphi_\alpha\left(q'(x,\omega)\right) = E_P\left[f_x(\omega)q'(x,\omega)\right]$$

- φ_α is known as Conditional Value-at-Risk in the mathematical finance literature.

The Upshot

- Our original problem becomes:

$$\min_{x \in X} c(x) + \varphi_\alpha\left(q'(x,\omega)\right)$$

$$= \min_{x \in X, \beta \in R} \left\{ c(x) + \beta + \frac{1}{1-\alpha} E_P\left[\left(q'(x,\omega) - \beta\right)^+\right] \right\}$$

- But this just looks like the one we have at the beginning of the talk!

The Upshot

- We can prove a sampling theorem similar to that of Charikar et al. under the new robust setting.
  - Sample complexity for discrete X:

$$poly\left(\log|X|, \lambda, \varepsilon^{-1}, \log\delta^{-1}, (1-\alpha)^{-1}\right)$$

- Also, we have a convex program (jointly in (x, β)) if c, q, X are convex.

How About SSC?

- Observe that φ_α is positive homogeneous and monotone, i.e. $\varphi_\alpha(cU) = c\varphi_\alpha(U)$ and $\varphi_\alpha(U) \le \varphi_\alpha(V)$ if $U \le V$.
- Thus, we can proceed as before.
  - Solve the convex relaxation of the sampled problem.
  - Examine the first-stage variables to determine the rounding.

Conclusion and Open Problems

- We have shown:
  - how to incorporate robustness in stochastic programs;
  - sampling is possible even with robustness!
- A great way to make simple problems hard (and hence create open problems)!
- Find more applications.

Thank You!