Convex Optimization II

Lecture 13: Stochastic Programming: Optimization under Uncertainty

Hamed Shah-Mansouri

Department of Electrical Engineering

Sharif University of Technology

1400-2

OUTLINE

- Introduction to stochastic programming
- Two-stage stochastic programming
- Scenario construction and Monte Carlo sampling methods
- Certainty-equivalent problem
- Multi-stage stochastic programming

MOTIVATIONS

- Real world problems usually include parameters which are unknown at the time a decision should be made.
- So far, we have studied deterministic problems which are formulated with known parameters.
- Stochastic programming is an approach for modelling optimization problems that involve uncertainty.
- In the absence of data from future periods, it may be difficult to arrive at an optimal decision.

[1] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modelling and Theory.* Society for Industrial and Applied Mathematics (SIAM), 2nd Ed., 2014.

[2] A. Shapiro and A. Philpott, "A Tutorial on Stochastic Programming", [Online]. Available.
http://www.isye.gatech.edu/people/faculty/Alex_Shapiro/TutorialSP.pdf

The Newsvendor Problem

- A company has to decide an order quantity x of a certain product to satisfy demand d. The cost of ordering is $c > 0$ per unit.
- If $d > x$, then a back order penalty of $b \geq 0$ per unit is incurred with a total cost of $b(d - x)$. Usually, $b > c$.
- If $d < x$, then a holding cost of $h(x - d)$ is incurred.
- The total cost is then

$$G(x, d) = cx + b[d - x]_+ + h[x - d]_+ \tag{1}$$

where $[y]_+ = \max\{y, 0\}$.

The Newsvendor Problem

- The objective is to minimize the total cost $G(x, d)$.
- If the demand is known, the corresponding optimization problem can be formulated in the form

$$\underset{x \geq 0}{\text{minimize}}\ G(x, d)$$

- We can rewrite

$$G(x, d) = \max\left\{(c-b)x + bd, (c+h)x - hd\right\},$$

which is piecewise linear with a minimum attained at $x^* = d$. No surprise!

- That is, if the demand d is known, then the best decision is to order exactly the demand quantity d.

The Newsvendor Problem

Question: What if the ordering decision should be made before a realization of the demand becomes known?

- One possible way is to view the demand as a random variable D.
- We assume that the probability distribution of D is known a *priori*.
- Then, alternatively, we can optimize the total cost on average.

$$\underset{x \geq 0}{\text{minimize}}\ \mathbb{E}G(x, D) \tag{2}$$

- A solution of problem (2) will be optimal on average given the law of large numbers.

Recourse Action

- The above problem gives a simple example of a recourse action.
- At the *first stage*, before a realization of the demand D is known, one has to make a decision about ordering quantity x.
- At the *second stage*, after demand D becomes known, it may happen that $d > x$. In that case, the company can meet demand by taking the recourse action of ordering the required quantity $d - x$ at a penalty cost of $b > c$.

Basic Stochastic Programming Problem

$$\begin{aligned} \underset{x}{\text{minimize}}\ & F_0(x) = \mathbb{E} f_0(x, D) \\ \text{subject to}\ & F_i(x) = \mathbb{E} f_i(x, D) \leq 0, \quad i = 1, \ldots, m \end{aligned}$$

- Problem data are f_i and distribution of D
- If $f_i(x, D)$ are convex in x for each D, $F_i(x)$ are convex, so is the stochastic programming problem.

Question: How to solve the above optimization problem?

How to Solve?

- Closed–form solutions
  - If the distribution of d is known and F_i have analytical expressions.
- Approximate solutions
  - Scenario construction
  - Monte Carlo sampling methods
  - Certainty-equivalent problem
  - Chance-constrained optimization methods
  - Robust optimization

Closed–form Solutions

- If the variable D is continuous and its distribution is known, the stochastic programming problem turns into a deterministic problem involving an integral term.

- Consider random variable D has a finitely supported distribution, i.e., it takes values $d_1, \ldots, d_K$, called scenarios, with probabilities $p_1, \ldots, p_K$. One can write

$$\mathbb{E}G(x, D) = \sum_{k=1}^{K} p_k G(x, d_k)$$

Closed–form Solutions

In the case of one scenario d taken with probability one,

$$\underset{x \geq 0}{\text{minimize}}\ G(x,d) = \max\{(c-b)x + bd, (c+h)x - hd\}$$

which is equivalent to the following LP problem.

$$\begin{aligned} \underset{x,t}{\text{minimize}}\ & t \\ \text{subject to}\ & t \geq (c-b)x + bd, \\ & t \geq (c+h)x - hd, \\ & x \geq 0. \end{aligned}$$

Closed–form Solutions

Now, with finitely many scenarios

$$
\begin{aligned}
\underset{x,\mathbf{t}}{\text{minimize}} \quad & \sum_{k=1} p_k t_k \\
\text{subject to} \quad & t_k \geq (c-b)x + bd_k, \quad k = 1, \ldots, K \\
& t_k \geq (c+h)x - hd_k, \quad k = 1, \ldots, K \\
& x \geq 0.
\end{aligned}
$$

The tractability of LP problems makes approximation by scenarios an attractive approach for attacking the stochastic programming problems.

Scenario Construction

Approximating an unknown distribution with scenarios.

Looks promising but raises several questions:

1. First, how should we approximate the random variable by one with a finitely-supported probability distribution?
2. Second, how should we solve the resulting (approximate) optimization problem?
3. Third, how can we measure the quality of the obtained solutions with respect to the true optimum?

Two-stage stochastic programming

Decisions (not necessarily optimal) should be based on data available at the time the decisions are made and should not depend on future observations.

The goal is to minimize the cost. In case of no uncertainty, we could have

$$\begin{aligned} &\underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}}\ \mathbf{c}^T\mathbf{x} \\ &\text{subject to } T\mathbf{x} \leq s. \end{aligned}$$

The Basic Concept

- At the first stage, we have to make a decision $\mathbf{x}$ before the realization of the uncertain data ζ is known.
- At the second stage, after a realization of ζ becomes available, we optimize our behavior by solving an appropriate optimization problem.

Two-stage Stochastic Programming

The classical two-stage linear stochastic programming problems can be formulated as:

$$\text{1st Stage: } \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}}\ \mathbf{c}^T\mathbf{x} + \mathbb{E}[Q(x,\zeta)]$$

where $Q(x,\zeta)$ is the optimal value of the second-stage problem:

$$\begin{aligned} \text{2nd Stage: } & \underset{\mathbf{y}}{\text{minimize}}\ \mathbf{q}^T\mathbf{y} \\ & \text{subject to } T\mathbf{x} + W\mathbf{y} \le s. \end{aligned}$$

- $\mathbf{x} \in \mathbb{R}^n$ is the first-stage decision vector,
- $\mathbf{y} \in \mathbb{R}^m$ is the second-stage decision vector,
- $\zeta = (q, T, W, s)$ contains the data of the second-stage problem.
- T is referred as tender, W is called the recourse matrix.

The Solution

- At the first stage, we minimize the cost $\mathbf{c}^T\mathbf{x}$ of the first-stage decision plus the expected cost of the (optimal) second-stage decision.
- At the second stage, we solve an optimization problem which describes our supposedly optimal behavior when the uncertain data is revealed. The term $W\mathbf{y}$ compensates for a possible inconsistency of the system $T\mathbf{x} \le s$ and $\mathbf{q}^T\mathbf{y}$ is the cost of the recourse action.

Two-stage stochastic programming

How to Solve?

If random vector ζ has a finite number of possible realizations, i.e., scenarios, say $\zeta_1, \ldots, \zeta_K$, with probabilities $p_1, \ldots, p_K$. Then, the expectation can be written as

$$\mathbb{E}[Q(x, \zeta)] = \sum_{k=1}^{K} p_k Q(x, \zeta_k)$$

The two-stage problem would be

$$\begin{aligned} &\underset{\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_K}{\text{minimize}} \ \mathbf{c}^T \mathbf{x} + \sum_{k=1}^{K} p_k q_k^T \mathbf{y}_k \\ &\text{subject to } T_k \mathbf{x} + W_k \mathbf{y}_k \leq s_k, \quad k = 1, \ldots, K. \end{aligned}$$

Two-stage stochastic programming

However, when ζ has an infinite (or very large) number of possible realizations, the standard approach is then to represent this distribution approximately by scenarios.

Some questions remain unanswered.

1. First, how should we approximate the random variable by one with a finitely-supported probability distribution?
2. ~~Second, how should we solve the resulting (approximate) optimization problem?~~
3. Third, how can we measure the quality of the obtained solutions with respect to the true optimum?

SCENARIO CONSTRUCTION

- In practice, it might be possible to construct scenarios by collecting experts' opinions on the future.
- If we choose a few number of scenarios, the accuracy may not be adequate.
- If we choose many scenarios, it may be computationally intractable.
- What is a reasonable number of scenarios to choose so that an obtained solution will be close to the true optimum of the original problem?

Monte Carlo Sampling Method

A possible approach to deal with the aforementioned two contradictory goals is by randomization. That is, scenarios could be generated by Monte-Carlo sampling techniques.

Randomly generate N replications of the random variable $\zeta^1, \dots, \zeta^N$ usually with the same distribution and independent identically distributed (iid).

Given a sample, we can approximate the expectation function $q(x) = \mathbb{E}[Q(x, \zeta)]$ by the average

$$\tilde{q}_N(x) = \frac{1}{N} \sum_{j=1}^{N} Q(x, \zeta^j)$$

Theory says, with some technical conditions, when $N \to \infty$, $x^*_{mcs} \to x^*$

Certainty-equivalent Problem

Idea: Ignore parameter variation

Certainty-equivalent (a.k.a. mean field) problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \; & F_0(x) = f_0(x, \mathbb{E}d) \\ \text{subject to} \; & F_i(x) = f_i(x, \mathbb{E}d) \leq 0, \quad i = 1, \ldots, m \end{aligned}$$

If f_i are convex in d for each x, then

$$f_i(x, \mathbb{E}d) \leq \mathbb{E} f_i(x, d)$$

Thus, the certainty-equivalent problem provides a lower bound on the optimal value of the stochastic problem.

Multi-stage Stochastic Programming

So far, we made a (supposedly optimal) decision at one point in time, while accounting for possible recourse actions after all uncertainty has been resolved.

There are many situations where decisions should be made sequentially at certain periods of time based on information available at each time period.

The Newsvendor Problem in Finite Horizon

The company has a planning horizon of T periods of time.

Let's view demand D_t as a random process indexed by time $t = 1, \ldots, T$.

- In the beginning, at $t = 1$, there is (known) inventory level y_1.
- At each period $t = 1, \ldots, T$, the company first observes the current inventory level y_t and then fills up the inventory level to x_t (order quantity $x_t - y_t \geq 0$).
- At the beginning of period $t + 1$, once demand d_t is realized, the inventory level becomes $y_{t+1} = x_t - d_t$.

The total cost incurred in period t is

$$c_t(x_t - y_t) + b_t[d_t - x_t]_+ + h_t[x_t - d_t]_+.$$

where c_t, b_t, and h_t are the ordering cost, backorder penalty cost and holding cost per unit, respectively, at time t.

Multi-stage Stochastic Programming

The objective is to minimize the expected value of the total cost over the planning horizon.

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \quad & \sum_{t=1}^{T} \mathbb{E}\left\{c_t(x_t - y_t) + b_t[D_t - x_t]_+ + h_t[x_t - D_t]_+\right\} \\
\text{subject to} \quad & y_{t+1} = x_t - D_t, \quad t = 1, \ldots, T-1 \\
& x_t \geq y_t, \qquad\qquad\quad t = 1, \ldots, T
\end{aligned}
$$

Multi-stage Stochastic Programming

Let $D_{[t]} = (D_1, \ldots, D_t)$ denote the history of the demand process up to time t, and $d_{[t]} = (d_1, \ldots, d_t)$ denote its particular realization.

At each t, our decision about x_t should depend only on information available at the time of the decision, i.e., on an observed realization $d_{[t-1]}$ of the demand process, and not on future observations.

Assume that the probability distribution of the demand process is known. That is, the conditional probability distribution of D_t, given $D_{[t-1]} = d_{[t-1]}$, is known.

Multi-stage Stochastic Programming

At the last stage $t = T$, we need to solve the problem:

$$\underset{x_T}{\text{minimize}}\ c_T(x_T - y_T) + \mathbb{E}\left\{b_T[D_T - x_T]_+ + h_T[x_T - D_T]_+ \mid D_{[T-1]} = d_{[T-1]}\right\}$$
$$\text{subject to } x_T \geq y_T.$$

The above expectation is taken conditional on realization $d_{[T-1]}$ prior to the considered time T.

Let denote the optimal value of the above problem as $V_T(y_T, d_{[T-1]})$.

At stage $t = T-1$, we solve the problem

روز 29

هزینه بهینه روزهای 29 و 30

اتفاقات تا 28

$$
\begin{aligned}
V_{T-1}(y_{T-1}, d_{[T-2]}) = \underset{x_{T-1}}{\text{minimize}}\ & c_{T-1}(x_{T-1} - y_{T-1}) \\
& + \mathbb{E}\{b_{T-1}[D_{T-1} - x_{T-1}]_+ \\
& + h_{T-1}[x_{T-1} - D_{T-1}]_+ \\
& + V_T(x_{T-1} - D_{T-1}, D_{[T-1]}) \mid D_{[T-2]} = d_{[T-2]}\} \\
\text{subject to}\ & x_{T-1} \geq y_{T-1}.
\end{aligned}
$$

روز 30

R.V

And so on, for stage $t = 1$, we reach

هزینه 2 to 30

$$
\begin{aligned}
V_1(y_1) = \underset{x_1}{\text{minimize}}\ & c_1(x_1 - y_1) \\
& + \mathbb{E}\{b_1[D_1 - x_1]_+ + h_1[x_1 - D_1]_+ + V_2(x_1 - D_1, D_1) \\
\text{subject to}\ & x_1 \geq y_1.
\end{aligned}
$$

$\to x_1, d_1 \longrightarrow y_1$

Multi-stage Stochastic Programming

For a generally distributed demand process, this could be very difficult or even impossible to solve the above set of problems.

The situation is simplified dramatically if we assume that the process D_t is *stagewise independent*.

Thus, value functions $V_t(y_t)$ do not depend on demand realizations and become functions of the respective univariate variables y_t only.

So, by using y_t and the (one-dimensional) distribution of D_t, these value functions can be calculated with a high accuracy in a *recursive way*.

Summary

- In optimization problems under uncertainty, a decision has to be made before the random parameters are known.
- Stochastic programming problems consider the expected value of the objective function and constraints.
- A popular technique to deal with uncertainty is two-stage stochastic programming.
- It is usually difficult to solve stochastic programming problems.
- Several approximation techniques are proposed to tackle the complexity of stochastic programming problems.