Stochastic Programming

Theory and Algorithms

Sirong Luo

Faculty of Statistics and Data Science
Shanghai University of Finance and Economics

Table of Contents

Examples of Stochastic Optimization Models

- The next three examples inherently involve making decisions under uncertainty.

  Example 10.1 (The newsvendor problem) A vendor purchases a particular commodity to satisfy some demand that occurs later over some time period. The demand D is random. The per-unit ordering cost, back-ordering cost, and holding costs are known to be c, p, and h, respectively. The total cost incured by the vendor if he purchases x units and the demand tums out to be D is

$$F(x, D) = c \cdot x + p \cdot \max(D - x, 0) + h \cdot \max(x - D, 0)$$

  The problem is to decide the order quantity x that minimizes the expected total cost $\mathbb{E}[F(x, D)]$.

Examples of Stochastic Optimization Models

- **Example 10.2 (Utility-based optimization)** An investor with endowment W_0 needs to decide how to invest this initial capital over a planning horizon. The investor's preferences for her final wealth W are expressed via a concave utility function $U(W)$. Assume $\mathbf{r}$ is the vector of random returns on the assets that the investor can purchase over the planning horizon. The investor wishes to choose a portfolio $\mathrm{x} \in \mathcal{X}$ that maximizes the expected utility $\mathbb{E}[U(W)]$ of her final wealth W :

$$W = W_0 \left(1 + \mathbf{r}^\top \mathbf{x}\right)$$

Examples of Stochastic Optimization Models

- **Example 10.3 (Optimal consumption and investment)** An individual may consume some portion C_0 of her initial endowment W_0 now and invest the remaining capital $W_0 - C_0$ for consumption at a future time. Assume $\mathbf{r}$ is the vector of random returns on the assets in which she can invest her remaining capital $W_0 - C_0$. Investing in a portfolio $\mathbf{x}$ will thus produce a random wealth $W = (W_0 - C_0)\left(1 + \mathbf{r}^\top \mathbf{x}\right)$. What should her consumption C_0 and investment decisions $\mathrm{x} \in \mathcal{X}$ be to maximize her total expected utility

$$U_0\left(C_0\right) + \mathbb{E}\left[U_1(W)\right]$$

Two-Stage Stochastic Optimization

- Consider the following generic type of optimization problem under uncertainty. At time 0 we need to make a set of decisions x subject to some constraint set $\mathcal{X}$. Between time 0 and time 1 a random outcome ω is revealed. Our goal is to choose $\mathbf{x}$ to minimize the expectation of some objective function $F(\mathbf{x}, \omega)$ that depends both on $\mathbf{x}$ as well as on the random outcome ω. This generic stochastic optimization problem has the following formal formulation:

$$\begin{aligned} \min_{\mathbf{x}} \mathbb{E}(F(\mathbf{x}, \omega)) \\ \mathbf{x} \in \mathcal{X} \end{aligned} \tag{10.1}$$

  The particular form of $F(\mathrm{x}, \omega)$ may define various types of problems, as we saw in Examples 10.1, 10.2, and 10.3. The function $F(\mathrm{x}, \omega)$ could be more involved: In a problem with recourse the function $F(\mathrm{x}, \omega)$ depends on decisions that can be made after the uncertainty ω is revealed.

- The particular form of $F(\mathbf{x}, \omega)$ may define various types of problems, as we saw in Examples 10.1, 10.2, and 10.3. The function $F(\mathbf{x}, \omega)$ could be more involved: In a problem with recourse the function $F(\mathbf{x}, \omega)$ depends on decisions that can be made after the uncertainty ω is revealed.

- Stochastic optimization with recourse is a refinement of the generic formulation (10.1). In this class of problems a first set of decisions $\mathbf{x}$ must be made here and now at time 0 . Between time 0 and time 1 a random outcome ω occurs. Then at time 1 we have the opportunity to make a new round of wait-and-see decisions $\mathbf{y}(\omega)$ after the random ω is revealed. This leads to a two-stage stochastic optimization with recourse problem formally stated as follows:

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) + \mathbb{E}[Q(\mathbf{x},\omega)] \\ \mathbf{x} \in \mathcal{X} \end{aligned} \tag{10.2}$$

- The recourse term $Q(\mathbf{x}, \omega)$ depends on the initial set of decisions $\mathbf{x}$ and on the random outcome ω, and it is of the form

$$Q(\mathbf{x}, \omega) := \min_{\mathbf{y}(\omega)} \quad \begin{array}{l} g(\mathbf{y}(\omega), \omega) \\ \mathbf{y}(\omega) \in \mathcal{Y}(\mathbf{x}, \omega) \end{array}$$

  The set of decisions $\mathbf{y}(\omega)$ are the recourse decisions. They are adaptive to the random outcome ω. This means that unlike $\mathbf{x}$ they are allowed to depend on ω.

Two-Stage Stochastic Optimization

- Example 10.4 (Newsvendor problem revisited) In this case the total cost is

$$F(x, D) = c \cdot x + p \cdot \max(D - x, 0) + h \cdot \max(x - D, 0).$$

We want to solve

$$\begin{aligned} \min_{x \geq 0} \mathbb{E}[F(x, D)] &= \min_{x \geq 0}(c \cdot x + \mathbb{E}[p \cdot \max(D - x, 0) + h \cdot \max(x - D, 0)]) \\ &= \min_{x \geq 0}(c \cdot x + \mathbb{E}[Q(x, D)]) \end{aligned}$$

where the recourse term $Q(x, D)$ is

$$\begin{aligned} Q(x, D) := \min_{y,z} \quad & py + hz \\ \text{s.t.} \quad & y \geq D - x \\ & z \geq x - D \\ & y, z \geq 0. \end{aligned}$$

Two-Stage Stochastic Optimization

- Note that here the recourse decisions y and z are easy to compute once the demand D and the number of units purchased x are known, namely $y = (D - x)^+$ and $z = (x - D)^+$.
  Sometimes it is preferable to consider a more general model obtained by replacing the objective $\mathbb{E}(F(\mathbf{x}, \omega))$ in (10.1) with $\varrho(F(\mathbf{x}, \omega))$ where $\varrho(\cdot)$ is a realvalued function. In particular, it is common to let $\varrho(\cdot)$ be a risk measure as illustrated in the following example. We formally define and discuss risk measures in more detail in Chapter 11.

Two-Stage Stochastic Optimization

- Example 10.5 (Mean-variance revisited) Let $\mathbf{r}$ denote the vector of random asset returns in an investment universe and let $\boldsymbol{\mu}$ and $\mathbf{V}$ denote respectively the expected value and covariance matrix of $\mathbf{r}$. The classic mean-variance model

$$\begin{array}{ll} \min_{\mathbf{x}} & \frac{1}{2}\gamma \cdot \mathbf{x}^\top \mathbf{V}\mathbf{x} - \boldsymbol{\mu}^\top \mathbf{x} \\ & \mathbf{x} \in \mathcal{X} \end{array}$$

can be written as the stochastic optimization problem

$$\begin{array}{c} \min_{\mathbf{x}} \varrho\left(\mathbf{r}^\top \mathbf{x}\right) \\ \mathbf{x} \in \mathcal{X} \end{array}$$

for the risk measure ϱ defined by

$$\varrho(Z) = \frac{1}{2}\gamma \cdot \sigma^2(Z) - \mathbb{E}(Z).$$

Linear Two-Stage Stochastic Programming

- A linear two-stage stochastic program is a problem of the form

$$\begin{array}{ll} \min_{\mathbf{x}} & \mathbf{c}^\top \mathbf{x} + \mathbb{E}[Q(\mathbf{x},\omega)] \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{array}$$

where the recourse term $Q(\mathbf{x},\omega)$ is the value of another linear program:

$$\begin{array}{rl} Q(\mathbf{x},\omega) := \min_{\mathbf{y}} & \mathbf{q}(\omega)^\top \mathbf{y}(\omega) \\ \text{s.t.} & \mathbf{T}(\omega)\mathbf{x} + \mathbf{W}(\omega)\mathbf{y}(\omega) = \mathbf{h}(\omega) \\ & \mathbf{y}(\omega) \geq 0 \end{array}$$

- Here the parameters $\mathbf{q}(\omega), \mathbf{T}(\omega), \mathbf{W}(\omega), \mathbf{h}(\omega)$ are random, and ω represents a random outcome $\omega \in \Omega$ that is revealed between stage 0 and stage 1 . It is customary and convenient to think of ω itself as the array of random parameters $\omega = (\mathbf{q}, \mathbf{T}, \mathbf{W}, \mathbf{h})$. The vector $\mathbf{x}$ represents the first-stage decisions. These must be made without knowing the random draw ω. The vector $\mathbf{y}(\omega)$ denotes the secondstage decisions. These may depend on the random draw ω.

Linear Two-Stage Stochastic Programming

- To ease notation, this type of problem is often written in the following form:

$$\begin{aligned} \min \mathbb{E}\left[\mathbf{c}^\top \mathbf{x} + \mathbf{q}^\top \mathbf{y}\right] & \\ \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} & \\ \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} = \mathbf{d} & \\ \mathbf{x} \geq \mathbf{0} & \\ \mathbf{y} \geq \mathbf{0} & \end{aligned} \tag{10.3}$$

but we should keep in mind that the tuple of uncertain parameters $\omega = (\mathbf{q}, \mathbf{T}, \mathbf{W}, \mathbf{h})$ is revealed between time 0 and time 1 and the recourse variables $\mathbf{y}$ may depend on this outcome.

Scenario Optimization

- Scenario optimization is a computational approach to stochastic optimization. The gist of this approach is to assume a discrete distribution for the random outcome. More precisely, assume the set of possible random outcomes is a discrete probability space $\Omega = \{\omega_1, \ldots, \omega_S\}$, with probability distribution $p_k = \mathbb{P}(\omega_k), k = 1, \ldots, S$. The elements in Ω are the possible realizations or scenarios

$$\omega_k = (\mathbf{q}_k, \mathbf{T}_k, \mathbf{W}_k, \mathbf{h}_k)$$

of the stochastic components of the model.

- Under this assumption, the stochastic optimization problem (10.3) can be written as the following deterministic equivalent:

$$\begin{aligned}
\min_{\mathbf{x},\mathbf{y}_k} \ & \mathbf{c}^\top \mathbf{x} + \sum_{k=1}^{S} p_k \left(\mathbf{q}_k^\top \mathbf{y}_k \right) \\
\text{s.t.} \ & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{T}_k \mathbf{x} + \mathbf{W}_k \mathbf{y}_k = \mathbf{h}_k \qquad \text{for } k = 1, \ldots, S \\
& \mathbf{x} \geq \mathbf{0} \\
& \mathbf{y}_k \geq \mathbf{0} \qquad \text{for } k = 1, \ldots, S
\end{aligned} \tag{10.4}$$

Scenario Optimization

- The deterministic equivalent problem has S copies of the second-stage decision variables and hence can be significantly larger than the original problem before we considered the uncertainty of the parameters. Fortunately, the constraint matrix has a very special sparsity structure that can be exploited as we explain in Section 10.5 below.

*The L-Shaped Method

- The constraint matrix of (10.4) has the following form:

$$\begin{bmatrix} \mathbf{A} & & & \\ \mathbf{T}_1 & \mathbf{W}_1 & & \\ \vdots & & \ddots & \\ \mathbf{T}_S & & & \mathbf{W}_S \end{bmatrix}$$

Observe that the blocks $\mathbf{W}_1, \ldots, \mathbf{W}_S$ of the constraint matrix are only interrelated through the blocks $\mathbf{T}_1, \ldots, \mathbf{T}_S$ which correspond to the first-stage decisions. In other words, once the first-stage decisions $\mathbf{x}$ have been fixed, (10.4) decomposes into S independent linear programs. The Benders decomposition method is an algorithm that takes advantage of this type of structure. This method is also called the L-shaped method in the stochastic programming literature.

*The L-Shaped Method

- The constraint matrix of (10.4) has the following form:

$$\begin{bmatrix} \mathbf{A} & & & \\ \mathbf{T}_1 & \mathbf{W}_1 & & \\ \vdots & & \ddots & \\ \mathbf{T}_S & & & \mathbf{W}_S \end{bmatrix}$$

  The idea behind this method is to solve a "master problem" involving only the variables $\mathbf{x}$ and a series of independent "recourse problems" each involving a different vector of variables $\mathbf{y}_k$. The master problem and recourse problems are linear programs. The size of these linear programs is much smaller than the size of the full model (10.4). The recourse problems are solved for a given vector $\mathbf{x}$ and their solutions are used to generate inequalities that are added to the master problem. Solving the new master problem produces a new $\mathbf{x}$ and the process is repeated.

*The L-Shaped Method

- More specifically, let us write (10.4) as

$$\begin{aligned} \min_{\mathbf{x}} \ & \mathbf{c}^\top \mathbf{x} + P_1(\mathbf{x}) + \cdots + P_S(\mathbf{x}) \\ \text{s.t.} \ & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{10.5}$$

where

$$\begin{aligned} P_k(\mathbf{x}) = \min_{\mathbf{y}_k} \ & p_k \mathbf{q}_k^\top \mathbf{y}_k \\ \text{s.t.} \ & \mathbf{W}_k \mathbf{y}_k = \mathbf{h}_k - \mathbf{T}_k \mathbf{x} \\ & \mathbf{y}_k \geq \mathbf{0} \end{aligned} \tag{10.6}$$

*The L-Shaped Method

for $k = 1, \dots, S$. The dual of the linear program (10.6) is:

$$\begin{aligned} P_k(\mathbf{x}) = \max_{\mathbf{u}_k} & \left(\mathbf{h}_k - \mathbf{T}_k\mathbf{x}\right)^\top \mathbf{u}_k \\ \text{s.t.} \ & \mathbf{W}_k^\top \mathbf{u}_k \le p_k \mathbf{q}_k \end{aligned} \tag{10.7}$$

*The L-Shaped Method

- For simplicity, assume (10.7) is feasible, which is the case of interest in many applications. The recourse linear program (10.6) will be solved for a sequence of vectors $\mathbf{x}^i$, for $i = 0, 1, 2, \ldots$. The initial vector $\mathbf{x}^0$ can be obtained by solving

$$\begin{aligned} \min_{\mathbf{x}} & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \;\; & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{10.8}$$

  For a given vector $\mathbf{x}^i$, two possibilities can occur for the recourse linear program (10.6): either (10.6) has an optimal solution or it is infeasible.

*The L-Shaped Method

- If (10.6) has an optimal solution $\mathbf{y}_k^i$, and $\mathbf{u}_k^i$ is the corresponding optimal dual solution, then (10.7) implies that

$$P_k(\mathbf{x}) \geq \left(\mathbf{u}_k^i\right)^\top \left(\mathbf{T}_k\mathbf{x}^i - \mathbf{T}_k\mathbf{x}\right) + P_k\left(\mathbf{x}^i\right).$$

  This inequality, which is called an optimality cut, can be added to the current master linear program. Initially, the master linear program is just (10.8).

*The L-Shaped Method

- If (10.6) is infeasible, then the dual problem is unbounded. Let $\mathbf{u}_k^i$ be a direction where (10.7) is unbounded, that is, $\left(\mathbf{h}_k - \mathbf{T}_k\mathbf{x}^i\right)^\top \mathbf{u}_k^i > 0$ and $\mathbf{W}_k^\top \mathbf{u}_k^i \leq p_k\mathbf{q}_k$. Since we are only interested in first-stage decisions $\mathbf{x}$ that lead to feasible second-stage decisions $\mathbf{y}_k$, the following feasibility cut can be added to the current master linear program:

$$\left(\mathbf{u}_k^i\right)^\top (\mathbf{h}_k - \mathbf{T}_k\mathbf{x}) \leq 0.$$

*The L-Shaped Method

- After solving the recourse problems (10.6) for each k, we have the following upper bound on the optimal value of (10.4):

$$UB = \mathbf{c}^\top \mathbf{x}^i + P_1\left(\mathbf{x}^i\right) + \cdots + P_S\left(\mathbf{x}^i\right),$$

  where we set $P_k\left(\mathbf{x}^i\right) = +\infty$ if the corresponding recourse problem is infeasible.

*The L-Shaped Method

- Adding all the optimality and feasibility cuts found so far (for $j = 0, \ldots, i$) to the master linear program, we obtain:

$$\begin{array}{cl} \min_{\mathbf{x}, z_1, \ldots, z_S} & \mathbf{c}^\top \mathbf{x} + \sum_{k=1}^S z_k \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array}$$

  for some pairs (j, k) :

$$\left(\mathbf{u}_k^j\right)^\top \left(\mathbf{T}_k \mathbf{x}^j - \mathbf{T}_k \mathbf{x}\right) + P_k\left(\mathbf{x}^j\right) \leq z_k$$

  for the remaining pairs (j, k) :

$$\left(\mathbf{u}_k^j\right)^\top (\mathbf{h}_k - \mathbf{T}_k \mathbf{x}) \leq 0$$

$$\mathbf{x} \geq \mathbf{0}$$

*The L-Shaped Method

- Denoting by $\mathbf{x}^{i+1}, z_1^{i+1}, \ldots, z_S^{i+1}$ an optimal solution to this linear program, we get a lower bound on the optimal value of (10.4):

$$LB = \mathbf{c}^\top \mathbf{x}^{i+1} + z_1^{i+1} + \cdots + z_S^{i+1}$$

  The Benders decomposition method alternately solves the recourse problems (10.6) and the master linear program with new optimality and feasibility cuts added at each iteration until the gap between the upper bound UB and the lower bound LB falls below a given threshold. It can shown that $UB - LB$ converges to zero and indeed reaches zero after finitely many iterations. For details see Birge and Louveaux (1997, chapter 5).