OPER 732: Optimization under Uncertainty
Lecture 5: Robust Optimization Models

Department of Statistical Sciences and Operations Research
Virginia Commonwealth University

Sept. 18, 2017

Stochastic optimization and robust optimization

Critics of stochastic programming:

- No perfect knowledge about probability distribution:
  - Estimation errors may cause suboptimal solutions or even infeasible solutions
- Computationally challenging

Robust optimization (RO):

- Instead of a probability distribution, a RO is defined by an uncertainty set
- The robust counterpart of a linear program remains a linear program (of a similar size), so computationally tractable
- RO incorporates the decision makers' risk tolerance, and combine this information with the historical data to construct the uncertainty set

Outline

1. Static robust optimization models
2. Two-stage robust optimization models

Outline

1. Static robust optimization models
2. Two-stage robust optimization models

Single-stage (static) model

Setting:

- Make a decision before uncertainty reveals
- No recourse actions available

Assumption:

- The uncertain parameter lies in a certain interval, called the range forecast
- The range forecast is usually symmetric around the point estimate, called the nominal value of the parameter

Consider a general LP:

$$\min \{c^\top x \mid Ax \geq b, x \in X\},$$

where A is uncertain, and X is deterministic

Q: What if c is uncertain? What if b is uncertain?

Robust LP: static model

Suppose the possible values of each row i of A, A_i, is given by an uncertainty set $\mathcal{A}_i$, and the decision maker have to satisfy constraints under all possibilities:

$$
\begin{aligned}
&\min\ c^\top x \\
&\text{s.t. } a_i^\top x \geq b_i,\ \forall a_i \in \mathcal{A}_i,\ \forall i = 1, 2, \ldots, m \\
&\qquad x \in X,
\end{aligned}
$$

- The robust counterpart is harder to solve than the nominal problem
- But it is still tractable in some cases

Uncertainty set and budget of uncertainty

- Suppose each entry a_{ij} has a range forecast of $[\bar{a}_{ij} - \hat{a}_{ij}, \bar{a}_{ij} + \hat{a}_{ij}]$, and nominal value is $\bar{a}_{ij}$
- Scaled deviation $z_{ij} := \frac{a_{ij} - \bar{a}_{ij}}{\hat{a}_{ij}}$, so $z_{ij} \in [-1, 1]$
- So the total deviation, $\sum_{j=1}^{n} z_{ij}$, in theory could be any number between $-n$ and n
- But...the value $\sum_{j=1}^{n} z_{ij}$ will take a narrow range! Some goes up, some goes down...

Budget of uncertainty

$$\sum_{j=1}^{n} |z_{ij}| \leq \Gamma_i, \forall i$$

Γ_i: the maximum allowable amount of deviation from the nominal values

A reasonable way of modeling uncertainty

$$\sum_{j=1}^{n} |z_{ij}| \leq \Gamma_i, \forall i$$

- One extreme: if $\Gamma_i = 0$, then $z_{ij} = 0, \forall j$, so parameters $a_{ij} = \bar{a}_{ij}$, no protection against uncertainty at all!
- The other extreme: if $\Gamma_i = n$, then the i-th constraint of A is fully protected (probably over protected)
- For $\Gamma_i \in (0, n)$, then a reasonable trade-off between protection level and the degree of conservatism

---

Uncertainty set:

$$\mathcal{A}_i = \{(a_{ij}) \mid a_{ij} = \bar{a}_{ij} + \hat{a}_{ij} z_{ij}, z \in Z_i\},$$

where $Z_i = \{z_{ij} \mid \sum_{j=1}^{n} |z_{ij}| \leq \Gamma_i, |z_{ij}| \leq 1, \forall j\}$

Robust LP

The robust LP using $\mathcal{A}$ defined in previous slide becomes:

$$\min\ c^\top x$$
$$\text{s.t. } \bar{a}_i^\top x + \min_{z_{ij}\in Z_i} \sum_{j=1}^{n} \hat{a}_{ij} x_j z_{ij} \geq b_i, \forall i$$
$$x \in X,$$

An optimization problem inside another optimization problem, how to deal with it?

Reformulation

Using LP duality, we reformulate the robust LP as an LP:

$$
\begin{aligned}
\min\ & c^\top x \\
\text{s.t. } & \bar{a}_i^\top x - \Gamma_i p_i - \sum_{j=1}^{n} q_{ij} \geq b_i, \forall i \\
& p_i + q_{ij} \geq \hat{a}_{ij} y_j, \forall i, j \\
& -y_j \leq x_j \leq y_j, \forall j \\
& p_i, q_{ij} \geq 0, \forall i, j \\
& x \in X,
\end{aligned}
$$

- Original LP: m constraints, n variables
- Reformulated robust LP: $n + (n+1)m$ new variables, $n(m+2)$ new constraints

How to choose the budget of uncertainty

- Γ reflects the decision makers's risk tolerance
- Select a budget so that $Ax \geq b$ is satisfied "with high probability in practice", without any information on the distribution of random matrix A

Connection between Γ and probability of violation

Assume $\Gamma_i \geq 1 + \Phi^{-1}(1 - \epsilon_i)\sqrt{n}$, then constraint $a_i^\top x \geq b_i$ is violated with probability at most ϵ_i, when each a_{ij} obeys a symmetric distribution centered at $\bar{a}_{ij}$ with support $[\bar{a}_{ij} - \hat{a}_{ij}, \bar{a}_{ij} + \hat{a}_{ij}]$.

e.g., $n = 150$, $\epsilon_i = 0.05$, then $\Gamma_i \geq 21.1$. Just need to protect against $14.1\%(= 21.1/150)$ of the worst-case values

Example: portfolio investment

Suppose we have 150 assets, the return of asset i belong to interval $[r_i - s_i, r_i + s_i]$: $r_i = 1.15 + i \times 0.05/150$, and $s_i = 0.05/450 \times \sqrt{300 \times 151 \times i}$

- Decision based on point estimator r_i: choose 150!
- Decision based on the worst-case scenario: choose 1!
- Decision based on robust optimization with performance guarantee: choose every asset with fraction decreases from 4.33% to 0.36% as i increases

Robust LP:

$$\begin{aligned} \max_{p, q_i, x_i \geq 0} \ & \sum_{i=1}^{150} r_i x_i - \Gamma p - \sum_{i=1}^{150} q_i \\ \text{s.t.} \ & \sum_{i=1}^{150} x_i = 1 \\ & p + q_i \geq s_i x_i \end{aligned}$$

Outline

From two-stage stochastic programming model to two-stage robust optimization model

$$\begin{aligned} \min\ & c^\top x + \mathbb{E}[q(\xi)^\top y(\xi)] \\ \text{s.t. } & Ax = b \\ & T(\xi)x + Wy(\xi) = h(\xi) \\ & x \in \mathbb{R}_+^{n_1}, y(\xi) \in \mathbb{R}_+^{n_2} \end{aligned}$$

$(T(\xi), q(\xi), h(\xi))$ follows a probability distribution

$$\begin{aligned} \min\ & c^\top x + \theta \\ \text{s.t. } & Ax = b \\ & \theta \geq q(\xi)^\top y(\xi),\ \forall \xi \in \Xi \\ & T(\xi)x + Wy(\xi) = h(\xi),\ \forall \xi \in \Xi \\ & x \in \mathbb{R}_+^{n_1}, y(\xi) \in \mathbb{R}_+^{n_2},\ \forall \xi \in \Xi \end{aligned}$$

$(T(\xi), q(\xi), h(\xi))$ belongs to an uncertainty set indexed by $\xi \in \Xi$

$(x, \{y(\xi)\}_{\xi \in \Xi})$ is a policy ($y(\xi)$ is a function of ξ), rather than a solution. Also called two-stage adjustable robust optimization model.

Two-stage adjustable vs. non-adjustable RO

Two-stage non-adjustable RO:

$$\begin{aligned} \min\ & c^\top x + \theta \\ \text{s.t. } & Ax = b \\ & \theta \geq q(\xi)^\top y, \ \forall \xi \in \Xi \\ & T(\xi)x + Wy = h(\xi), \ \forall \xi \in \Xi \\ & x \in \mathbb{R}_+^{n_1}, y \in \mathbb{R}_+^{n_2} \end{aligned}$$

Two-stage adjustable RO:

$$\begin{aligned} \min\ & c^\top x + \theta \\ \text{s.t. } & Ax = b \\ & \theta \geq q(\xi)^\top y(\xi), \ \forall \xi \in \Xi \\ & T(\xi)x + Wy(\xi) = h(\xi), \ \forall \xi \in \Xi \\ & x \in \mathbb{R}_+^{n_1}, y(\xi) \in \mathbb{R}_+^{n_2}, \ \forall \xi \in \Xi \end{aligned}$$

- In some cases, e.g., constraint-wise uncertainty (uncertainty affects different constraints independently) with fixed second-stage cost q, the two models are equivalent
- In general, adjustable RO provides more flexible (less conservative) decision, but it is much harder (NP-hard)

Example

Consider an uncertain linear programming problem with a single equality constraint: $\alpha u + \beta v = 1$, where the uncertain data (α, β) can take values in the uncertainty set

$$Z = \{(\alpha, \beta) \mid \alpha \in [\frac{1}{2}, 1], \beta \in [\frac{1}{2}, 1]\}.$$

- Feasible region of the adjustable RO model?
- Feasible region of the static RO model?

A simple case where ARO is tractable

When the uncertainty set is given by a convex hull of a set of points:

$$\{T(\xi), h(\xi)\}_{\xi \in \Xi} = conv(\{(T_1, h_1), (T_2, h_2), \cdots, (T_N, h_N)\})$$

The following ARO formulation is a linear program:

$$\begin{aligned}
\min\ & c^\top x + \theta \\
\text{s.t. } & Ax = b \\
& \theta \geq q^\top y(\xi),\ \forall \xi \in \Xi \\
& T(\xi)x + Wy(\xi) = h(\xi),\ \forall \xi \in \Xi \\
& x \in \mathbb{R}_+^{n_1}, y(\xi) \in \mathbb{R}_+^{n_2},\ \forall \xi \in \Xi
\end{aligned}$$

Affine adaptability

Restrict the policy to be affine function of uncertain data ξ:

$$y(\xi) = p + Q\xi$$

Then ARO with affine adaptability can be written as:

$$\begin{aligned}
\min\ & c^\top x + \theta \\
\text{s.t. } & Ax = b \\
& \theta \geq q^\top (p + Q\xi),\ \forall \xi \in \Xi \\
& T(\xi)x + W(p + Q\xi) = h(\xi),\ \forall \xi \in \Xi \\
& x \in \mathbb{R}^{n_1}_+, p, Q \text{ free}
\end{aligned}$$

ARO is computationally tractable, if the uncertainty set Z is computationally tractable, i.e., for any vector z there is a tractable "separation oracle". E.g., polyhedral/ellipsoidal uncertainty sets.

Finite adaptability

Idea: select a finite number of contingency plans to incorporate the information revealed over time

- Partition the uncertainty set into K pieces
- Determine the best response in each piece.

Appealing features of this approach:

- It provides a hierarchy of adaptability
- It is able to incorporate integer second-stage variables and nonconvex uncertainty sets, while other approaches (e.g., affine adaptability) cannot.

Right-hand side uncertainty

Robust optimization with rhs uncertainty:

$$\begin{aligned} &\min\ c^\top x \\ &\text{s.t. } Ax \geq b,\ \forall b \in \mathcal{B} \\ &\quad x \in X \end{aligned}$$

Equivalent to a deterministic problem with:

$$Ax \geq \tilde{b}_0,\ \text{where } (\tilde{b}_0)_i = \max\{b_i \mid b \in \mathcal{B}\}$$

Idea of K-adaptable robust optimization:

- Cover $\mathcal{B}$ with a collection of sets $\{\mathcal{B}_k\}_{k \in K}$ so that $\mathcal{B} \subseteq \bigcup_{k \in K} \mathcal{B}_k$.
- Select a contingency plan x_k for each subset $\mathcal{B}_k$.

K-adaptable robust optimization

$$
\begin{aligned}
\min \quad & \max_{k=1,2,\ldots,K} c^\top x_k \\
\text{s.t. } & Ax_k \geq b, \ \forall b \in \mathcal{B}_k, \ \forall k = 1,2,\ldots,K \\
& x_k \in X, \ \forall k = 1,2,\ldots,K
\end{aligned}
$$

Let $\tilde{b}_k$ be defined as $(\tilde{b}_k)_i = \max\{b_i \mid b \in \mathcal{B}_k\}$, then it becomes:

$$
\begin{aligned}
\min \quad & \max_{k=1,2,\ldots,K} c^\top x_k \\
\text{s.t. } & Ax_k \geq \tilde{b}_k, \ \forall k = 1,2,\ldots,K \\
& x_k \in X, \ \forall k = 1,2,\ldots,K
\end{aligned}
$$

One can optimize the collection of sets $\{\mathcal{B}_k\}_{k \in K}$ to get the best K-adaptable solution $\rightarrow$ However, this is an NP-hard combinatorial optimization problem (even for $K = 2$)!

Example: Newsvendor problem with reorder

A manager must order two types of items before knowing the actual demand for these products.

- All demand must be met
- The unit ordering price is 1
- Once demand is realized, the missing items (if any) are reordered at the unit price of 2
- The uncertainty set for the demand is given by:

$$\{(d_1, d_2) \mid d_1 \geq 0, d_2 \geq 0, \frac{d_1}{2} + d_2 \leq 1\}$$

The decision-maker considers two contingency plans. Let $x_j, j = 1, 2$ be the amounts of product j ordered before demand is known, and y_{ij} be the amount of product j ordered in contingency plan $i, i = 1, 2$.

To implement the 2-adaptability approach, suppose the decision-maker selects a covering pair $(d, 1)$ and $(1, 1 - d/2)$ with $0 \leq d \leq 2$. What is the optimal d and the corresponding contingency plan?

The optimal solution is to select $d = 2/3, x = (2/3, 2/3)$ and $y_1 = (0, 1/3), y_2 = (1/3, 0)$, for an optimal cost of 2.