IE 521: Convex Optimization	**Spring 2017, UIUC**

Lecture 16: Applications in Robust Optimization – March 27

Instructor: Niao He *Scribe: Shuanglong Wang*

Courtesy warning: These notes do not necessarily cover everything discussed in the class. Please email TA (swang157@illinois.edu) if you find any typos or mistakes.

In this lecture, we cover the following topics

- Robust Linear Program
- Example: Robust Portfolio Selection
- Example: Robust Classification

16.1 Robust Linear Program

Consider the linear program

$$\min \left\{ c^T x : Ax \leq b \right\} \tag{LP}$$

Assume that the data (c, A, b) of the program are not known exactly and vary in a given uncertainty set $\mathcal{U}$. The goal is to find a robust solution that is

- feasible for all instances, i.e. $Ax \leq b, \forall (c, A, b) \in \mathcal{U}$
- optimal for the worst-case objective, i.e. $\sup \left\{ c^T x : (c, A, b) \in \mathcal{U} \right\}$

We call the following program the robust counterpart of (LP):

$$\min_x \left\{ \sup_{(c,A,b)\in\mathcal{U}} c^T x : Ax \leq b, \forall (c, A, b) \in \mathcal{U} \right\} \tag{RC}$$

or equivalently,

$$\min_{x,t} \{t \ : c^T x \leq t \text{ and } Ax \leq b \quad \forall (c, A, b) \in \mathcal{U}\}$$

Note that the robust counterpart is a semi-infinite convex optimization program with infinitely many linear inequality constraints. The structure of (RC) depends on the geometry of the uncertainty set $\mathcal{U}$.

Without loss of generality, let's consider the robust counterpart in the simple form:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & a_i^T x \leq b_i, \quad \forall a_i \in \mathcal{U}_i, i = 1, ..., m \end{aligned} \tag{RC}$$

where $\mathcal{U}_i \subseteq \mathbf{R}^n$ is the uncertainty set of a_i.

Polyhedral Uncertainty

Consider the situation where the uncertainty set is a polyhedron:

$$\mathcal{U}_i = \{a_i : D_i a_i \le d_i\} \quad i = 1, ..., m$$

The (RC) becomes

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & \max_{a_i \in \mathcal{U}_i} a_i^T x \le b_i \end{aligned}$$

By LP duality, we know that

$$\begin{aligned} \max_{a_i} \quad & a_i^T x \\ \text{s.t.} \quad & D_i a_i \le d_i \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} \min_{p_i} \quad & d_i^T p_i \\ \text{s.t.} \quad & D_i^T p_i = x \\ & p_i \ge 0 \end{aligned}$$

Hence, (RC) is equivalent to

$$\begin{aligned} \min_{x, p_1, \dots, p_m} \quad & c^T x \\ \text{s.t.} \quad & d_i^T p_i \le b_i, i = 1, ..., m \\ & D_i^T p_i = x, i = 1, ..., m \\ & p_i \ge 0, i = 1, ..., m \end{aligned}$$

which is a linear program.

Ellipsoidal Uncertainty

Consider the situation when the uncertainty set is an ellipsoid

$$\mathcal{U}_i = \{\bar{a}_i + P_i u : \|u\|_2 \le 1\}, \quad i = 1, ..., m$$

where $\bar{a}_i \in \mathbf{R}^n, P_i \in \mathbf{R}^{n \times n}$.

The (RC) becomes

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & \max_{a_i \in \mathcal{U}_i} a_i^T x \le b_i \end{aligned}$$

Note that

$$\max_{a_i \in \mathcal{U}_i} a_i^T x = \max_{\|u\|_2 \le 1} \bar{a}_i^T x + u^T P_i^T x = \bar{a}_i^T x + \max_{\|u\|_2 \le 1} u^T P_i^T x = \bar{a}_i^T x + \|P_i^T x\|_2$$

Hence, (RC) is equivalent to

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & \|P_i^T x\|_2 \le b_i - \bar{a}_i^T x \end{aligned}$$

which is an SOCP.

16.2 Examples

16.2.1 Example I: Robust Portfolio Selection

Consider a classical portfolio optimization problem with n assets. Let $x = (x_1, ..., x_n)$ be the portfolio vector.

1. Assume that returns $r_i, i = 1, ..., n$ are exactly known. To maximize the return leads to the optimization problem.

$$\begin{aligned} \max_x \quad & r^T x \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 1 \\ & x \ge 0 \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \max_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & r^T x \ge \gamma \\ & \sum_{i=1}^n x_i = 1, x \ge 0 \end{aligned}$$

   which leads to a highly non-robust solution.

2. Assume the returns are know within an ellipsoid:

$$\mathcal{U} = \left\{ \hat{r} + \rho \hat{\Sigma}^{1/2} u : \quad \|u\|_2 \le 1 \right\}$$

   where $\hat{r}$ and $\hat{\Sigma}$ are the empirical mean and covariance matrix.

   The robust portfolio problem is

$$\max_x \min_{r \in \mathcal{U}} \gamma^T x, \text{ s.t. } x \ge 0, \sum_{i=1}^n x_i = 1$$

   which is equivalent as an SOCP:

$$\max_x \; \hat{r}^T x - \rho \|\hat{\Sigma}^{1/2} x\|_2, \text{ s.t. } x \ge 0, \mathbf{1}^T x = 1$$

   This can be interpreted as a return-risk tradeoff.

3. Assume that returns are random variables with mean $\hat{r}$ and covariance $\hat{\Sigma}$. Consider the chance constrained program:

$$\begin{aligned} \max_{x,\gamma} \quad & \gamma \\ \text{s.t.} \quad & P(r^T x \geq \gamma) \geq 1 - \epsilon \\ & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0 \end{aligned}$$

Note that

$$\begin{aligned} & P(r^T x \geq \gamma) \geq 1 - \epsilon \\ \Leftrightarrow \quad & P(Z \geq \frac{\gamma - \hat{r}^T x}{\sqrt{x^T \hat{\Sigma} x}}) \geq 1 - \epsilon, \text{ where } Z \sim N(0,1) \\ \Leftrightarrow \quad & P(Z \leq \frac{\hat{r}^T x - \gamma}{\sqrt{x^T \hat{\Sigma} x}}) \geq 1 - \epsilon \\ \Leftrightarrow \quad & \frac{\hat{r}^T x - \gamma}{\sqrt{x^T \hat{\Sigma} x}} \geq \Phi^{-1}(1 - \epsilon) \\ \Leftrightarrow \quad & \hat{r}^T x - \Phi^{-1}(1 - \epsilon) \|\hat{\Sigma}^{1/2} x\|_2 \geq \Gamma \end{aligned}$$

Hence, the chance constrained program is equivalent to

$$\max_{x} \quad \hat{r}^T x - \Phi^{-1}(1 - \epsilon) \|\hat{\Sigma}^{1/2} x\|_2, \text{ s.t. } x \geq 0, \mathbf{1}^T x = 1$$

Remark Robust LP with ellipsoidal uncertainty set are closed related to chance constraints of a stochastic model.

16.2.2 Example 2: Robust Classification

Consider the support vector machine model for binary classification:

$$\begin{aligned} \min_{\omega,b,\epsilon} \quad & \sum_{i=1}^{m} \epsilon_i \\ \text{s.t.} \quad & y_i(\omega^T x_i + b) \geq 1 - \epsilon_i \\ & \epsilon \geq 0 \end{aligned} \tag{SVM}$$

where $(x_i, y_i), i = 1, ..., m$ are data points with $x_i \in \mathbf{R}^n, y_i \in \{\pm 1\}$.

1. Assume that the feature vector x_i are subject to spherical uncertainty

$$x_i \in X_i = \{\hat{x}_i + \rho u : \|u\|_2 \leq 1\}$$

The robust counterpart of (SVM) simplifies to a SOCP

$$\begin{aligned}
\min_{\omega,b,\epsilon} \quad & \sum_{i=1}^{m} \epsilon_i \\
\text{s.t.} \quad & y_i(\omega^T \hat{x}_i + b) - \rho\|\omega\|_2 \geq 1 - \epsilon_i \\
& \epsilon_i \geq 0 \\
\Leftrightarrow \min_{\omega,b} \quad & \sum_{i=1}^{m} \max\Big(1 - y_i(\omega^T x_i + b) + \rho\|\omega\|_2, 0\Big)
\end{aligned}$$

which is very similar to the classical SVM with ℓ_2-norm regularization:

$$\min_{\omega,b} \quad \sum_{i=1}^{m} \max\Big(1 - y_i(\omega^T x_i + b), 0\Big) + \lambda\|\omega\|_2^2$$

2. Assume that the feature vector x_i are subject to box uncertainty

$$x_i \in X_i = \{\hat{x}_i + \rho u : \|u\|_\infty \leq 1\}$$

THe robust counterpart of (SVM) becomes

$$\min_{\omega,b} \quad \sum_{i=1}^{m} \max\Big(1 - y_i(\omega^T x_i + b) + \rho\|\omega\|_1, 0\Big)$$

which is very similar to the classical SVM with ℓ_1-norm regularization.

Remark Robust optimization has a close connection to the regularization technique in machine learning.