Chapter IX: Robust Optimization

1 Introduction

In many real-world applications, optimization problems depend on some data which cannot be known with certainty. This arises in different situations:

(i) The data collection can be a source of errors. For example, when the sensors collecting the data lack of accuracy, or when the data is entered manually.

(ii) In some cases, the system to be optimized contains some data which is not known, and had to be *estimated* based on historical data.

(iii) Sometimes, the uncertain data is even intrisically stochastic, and comes from a probability distribution which is either known or estimated from historical data.

Robust optimization is an important subfield of optimization that deals with uncertainty in the data. The constraints and objective are assumed to belong to some *uncertainty set*, and robust optimization asks to protect against the uncertainty, by taking decisions that are optimal for the worst-case, and that are feasible for all realizations of the uncertainty. Formally, we assume the problem to solve has the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathcal{X}}{\textbf{minimize}} \quad & f(\boldsymbol{x},\boldsymbol{\theta}) \\ s.t. \quad & g(\boldsymbol{x},\boldsymbol{\theta}) \leq 0, \end{aligned} \tag{P}$$

where $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ is the decision variable and $\boldsymbol{\theta} \in \mathbb{R}^k$ is some uncertain parameter. Then, the *robust counterpart* of (P) over the uncertainty set $\Theta \subseteq \mathbb{R}^k$ is

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathcal{X}}{\textbf{minimize}} \quad & \left(\sup_{\boldsymbol{\theta}\in\Theta} \quad f(\boldsymbol{x},\boldsymbol{\theta})\right) \\ s.t. \quad & g(\boldsymbol{x},\boldsymbol{\theta}) \leq 0,\ \forall \boldsymbol{\theta} \in \Theta. \end{aligned} \tag{RP}$$

In the sitation (iii) evoked above, when the uncertain data is stochastic, robust optimization provides an alternative to the *Stochastic Programming* paradigm, where $\boldsymbol{\theta}$ is a random variable and we try to solve

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathcal{X}}{\textbf{minimize}} \quad & \mathbb{E}_{\theta}[f(\boldsymbol{x},\boldsymbol{\theta})] \\ s.t. \quad & g(\boldsymbol{x},\boldsymbol{\theta}) \leq 0 \text{ almost surely.} \end{aligned} \tag{SP}$$

The worst-case nature of the robust counterpart is often criticized, because it tends to produce *overconservatism*, that is, solutions that are robust to extreme scenarios, but may perform rather poorly for the average scenarios. However, robust optimization remains a very good option to handle uncertainty:

- First, it does not require distributional information on the unkown parameter $\boldsymbol{\theta}$, which may be hard to obtain in practice.
- Second, the robust problem (RP) is often much easier to solve than the stochastic program (SP).
- Third, when the uncertainty set Θ is chosen wisely (e.g., a reasonable confidence region of the random variable $\boldsymbol{\theta}$), the robust counterpart generally produces solutions which perform quite well *on average*, too.

2 Robust Linear Programming

We introduce the robust counterpart of a linear program in this section, and we will see in the next section that it can be reformulated as a compact conic optimization problem for a large class of uncertainty models. The basic problem we study is as follows:

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\textbf{minimize}} \quad & \big(\sup_{\boldsymbol{\theta}\in\Theta}\ \boldsymbol{c}(\boldsymbol{\theta})^T\boldsymbol{x}\big) \\ s.t. \quad & A(\boldsymbol{\theta})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{\theta}), \quad \forall \boldsymbol{\theta}\in\Theta, \end{aligned} \tag{1}$$

where $\boldsymbol{c}(\boldsymbol{\theta}) \in \mathbb{R}^n$, $\boldsymbol{b}(\boldsymbol{\theta}) \in \mathbb{R}^m$, and $A(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$.

By using the epigraph form, and by introducing an auxiliary variable, we can assume w.l.o.g. that the uncertainty only occurs in the constraint matrix. That is, we restrict our study to problems of the form

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathcal{X}}{\textbf{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ s.t. \quad & A(\boldsymbol{\theta})\boldsymbol{x} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\theta}\in\Theta. \end{aligned} \tag{2}$$

To see this, note that Problem (1) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n,\ t,u\in\mathbb{R}}{\textbf{minimize}} \quad & t \\ s.t. \quad & A(\boldsymbol{\theta})\boldsymbol{x} - \boldsymbol{b}(\boldsymbol{\theta})u \leq 0, \quad \forall \boldsymbol{\theta}\in\Theta, \\ & \boldsymbol{c}(\boldsymbol{\theta})^T\boldsymbol{x} - t \leq 0,\ \forall \boldsymbol{\theta}\in\Theta \\ & u = 1. \end{aligned} \tag{3}$$

which has the same form as (2) if we change the data as follows: $\boldsymbol{x}'^T = [\boldsymbol{x}^T, u, t]$, $\boldsymbol{c}'^T = [\boldsymbol{0}^T, 0, 1]^T$,

$$A'(\boldsymbol{\theta}) = \begin{bmatrix} A(\boldsymbol{\theta}) & -\boldsymbol{b}(\boldsymbol{\theta}) & \boldsymbol{0} \\ \boldsymbol{c}(\boldsymbol{\theta})^T & 0 & -1 \\ \boldsymbol{0}^T & 1 & 0 \\ \boldsymbol{0}^T & -1 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b}' = \begin{bmatrix} \boldsymbol{0} \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

Now, we consider Problem (2) again, and we denote by $\boldsymbol{a}_i(\boldsymbol{\theta})^T$ the ith row of $A(\boldsymbol{\theta})$. Since all constraints must be satisfied for all possible values of the uncertain parameter $\boldsymbol{\theta}$, the robust LP is equivalent to

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathcal{X}}{\textbf{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ s.t. \quad & \boldsymbol{a}_i(\boldsymbol{\theta})^T\boldsymbol{x} \leq b_i,\ \forall \boldsymbol{\theta}\in\Theta, \qquad \forall i\in[m]. \end{aligned} \tag{LP}$$

In other words, we can handle the uncertainty of each row *separately*. We say that the robust counterpart acts *constraint-wise*. We are thus left with the study of linear problems with constraints of the form

$$\boldsymbol{a}(\boldsymbol{\theta})^T\boldsymbol{x} \leq b, \quad \forall \boldsymbol{\theta}\in\Theta. \tag{4}$$

Such problems are known as semi-infinite linear optimization problems, because they involved a finite number of variables, but an infinite number of linear constraints (indexed by the continuous variable $\boldsymbol{\theta}\in\Theta$).

3 Robust counterpart of LPs under common uncertainty models

In this section, we will show how to reformulate semi-infinite constraints (4) as equivalent compact conic inequalities, for several uncertainty models. An *uncertainty model* gives the form of the function $\boldsymbol{a}(\boldsymbol{\theta})$ and of the uncertainty set Θ. Note that if we introduce the set $\mathcal{A} := \{\boldsymbol{a}(\boldsymbol{\theta})|\ \boldsymbol{\theta} \in \Theta\}$, the semi-infinite constraint (4) is equivalent to

$$\boldsymbol{a}^T\boldsymbol{x} \leq b, \quad \forall \boldsymbol{a} \in \mathcal{A}.$$

Hence an uncertainty model can be specified either by the set $\mathcal{A}$ or by the set Θ and the description of $\boldsymbol{a}(\cdot)$. In this lecture, we refer as *uncertainty set* both the sets Θ and $\mathcal{A}$. It is often more natural to introduce a parameter $\boldsymbol{\theta}$ which represents the uncertainty. Then, Θ represents the set of plausible *scenarios*, and $\mathcal{A}$ is the image of Θ by $\boldsymbol{a}(\cdot)$.

3.1 Polyhedral uncertainty

In the polyhedral uncertainty model, the vector $\boldsymbol{a}(\boldsymbol{\theta})$ is assumed to lie in a polyhedron, given by its vertices $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$. That is,

$$\mathcal{A} = \mathbf{conv}\,\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}.$$

Equivalently,

$$\boldsymbol{a}(\boldsymbol{\theta}) = \sum_{j\in[k]} \theta_j \boldsymbol{v}_j \quad \text{for some } \boldsymbol{\theta} \in \Theta := \{\boldsymbol{\theta} \geq \mathbf{0}|\ \mathbf{1}^T\boldsymbol{\theta} = 1\}.$$

Proposition 1. *In the polyhedral uncertainty model, the semi-infinite constraint* (4) *is equivalent to*

$$\boldsymbol{v}_j^T\boldsymbol{x} \leq b, \quad \forall j \in [k].$$

Proof. Let $\boldsymbol{x}$ satisfy the semi-infinite constraint, that is, $\boldsymbol{a}^T\boldsymbol{x} \leq b, \forall \boldsymbol{a} \in \mathcal{A}$. Then, we have $\boldsymbol{v}_j^T\boldsymbol{x} \leq b$ for all j, because $\boldsymbol{v}_j \in \mathcal{A}$. Assume conversely that $\boldsymbol{v}_j^T\boldsymbol{x} \leq b$, $\forall j \in [k]$. Consider an arbitrary convex combination of the $\boldsymbol{v}_j$'s: $\boldsymbol{a} = \sum_j \theta_j \boldsymbol{v}_j$ for some $\boldsymbol{\theta} \geq \mathbf{0}$ satisfying $\mathbf{1}^T\boldsymbol{\theta} = 1$. Multiplying each inequality by $\theta_j \geq 0$ and summing, we obtain

$$\sum_j \theta_j \boldsymbol{v}_j^T\boldsymbol{x} \leq \sum_j \theta_j b \iff \boldsymbol{x}^T(\underbrace{\sum_j \theta_j \boldsymbol{v}_j}_{=\boldsymbol{a}}) \leq b \underbrace{\sum_j \theta_j}_{=1}.$$

This shows: $\boldsymbol{a}^T\boldsymbol{x} \leq b, \forall \boldsymbol{a} \in \mathcal{A}$. □

3.2 Conic uncertainty

In the conic uncertainty model, $\boldsymbol{a}(\boldsymbol{\theta})$ is an affine function of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}$ belongs to a set defined by conic inequalities:

$$\mathcal{A} = \left\{ \underbrace{\bar{\boldsymbol{a}} + P\boldsymbol{\theta}}_{\boldsymbol{a}(\boldsymbol{\theta})} \ \middle|\ \ \boldsymbol{\theta} \in \Theta := \{\boldsymbol{\theta} \in \mathbb{R}^k : F\boldsymbol{\theta} \preceq_K \boldsymbol{h}\} \right\} \tag{5}$$

for some $\bar{\boldsymbol{a}} \in \mathbb{R}^n$, $P \in \mathbb{R}^{n\times k}$, $F \in \mathbb{R}^{\ell\times k}$, $\boldsymbol{h} \in \mathbb{R}^\ell$, and $K \subset \mathbb{R}^\ell$ a proper cone.

The following result shows that we can use the duality theory to reformulate the semi-infinite constraint (4) under the conic uncertainty model:

Theorem 2. *Assume that the relative interior of Θ is nonempty, or more generally that the conic inequality*

$$F\boldsymbol{\theta} \preceq_K \boldsymbol{h}$$

is essentially strictly feasible. Then, in the conic uncertainty model (5)*, the semi-infinite constraint* (4) *can be replaced by a system of conic inequalities which involves an additional variable $\boldsymbol{z} \in \mathbb{R}^\ell$:*

$$\big(\boldsymbol{a}(\boldsymbol{\theta})^T\boldsymbol{x} \leq b, \quad \forall \boldsymbol{\theta} \in \Theta\big) \iff \exists \boldsymbol{z} \in \mathbb{R}^\ell : \begin{cases} \bar{\boldsymbol{a}}^T\boldsymbol{x} + \boldsymbol{h}^T\boldsymbol{z} \leq b \\ F^T\boldsymbol{z} = P^T\boldsymbol{x} \\ \boldsymbol{z} \succeq_{K^*} \mathbf{0}. \end{cases}$$

Proof. We first rewrite the semi-infinite constraint as

$$\begin{aligned} &\sup_{\boldsymbol{\theta}\in\Theta} (\bar{\boldsymbol{a}} + P\boldsymbol{\theta})^T \boldsymbol{x} \leq b \\ \iff &\sup_{\boldsymbol{\theta}\in\Theta} \boldsymbol{\theta}^T P^T \boldsymbol{x} \leq b - \bar{\boldsymbol{a}}^T \boldsymbol{x} \end{aligned}$$

The expression on the LHS of the equality sign is a conic program with respect to the variable $\boldsymbol{\theta}$. This program is (essentially) strictly feasible by assumption, so strong duality holds, and

$$\begin{array}{llcll} \sup\limits_{\boldsymbol{\theta}} & \boldsymbol{\theta}^T P^T \boldsymbol{x} & = & \inf\limits_{\boldsymbol{z}} & \boldsymbol{z}^T \boldsymbol{h} \\ s.t. & F\boldsymbol{\theta} \preceq_K \boldsymbol{h} & & s.t. & F^T \boldsymbol{z} = P^T \boldsymbol{x}, \quad \boldsymbol{z} \succeq_{K^*} \boldsymbol{0}. \end{array}$$

(Note that $\boldsymbol{x}$ is treated as a constant here, because the decision variable is $\boldsymbol{\theta}$). We have thus shown that the semi-infinite constraint (4) is equivalent to

$$\exists \boldsymbol{z} \succeq_{K^*} \boldsymbol{0} : \begin{cases} F^T \boldsymbol{z} = P^T \boldsymbol{x} \\ \boldsymbol{z}^T \boldsymbol{h} \leq b - \bar{\boldsymbol{a}}^T \boldsymbol{x}. \end{cases}$$

□

The conic uncertainty model contains two very important special cases, which we present next.

3.2.1 Budgeted Uncertainty

In the budgeted uncertainty model, we specify a *nominal value* $\bar{\boldsymbol{a}}$ for the vector $\boldsymbol{a} \in \mathcal{A}$, as well as a vector of maximal deviations $\boldsymbol{\delta} \geq \boldsymbol{0}$, so that for all $i \in [n]$, the deviation from the nominal scenario satisfies:

$$|a_i - \bar{a}_i| \leq \delta_i.$$

In addition, we give a scalar $\Gamma \in [0, n]$ so that the sum of standardized deviations is $\leq \Gamma$:

$$\sum_{i=1}^{n} \frac{|a_i - \bar{a}_i|}{\delta_i} \leq \Gamma$$

If we denote by $\boldsymbol{\theta}$ the vector of standardized deviations, the budgeted uncertainty model is thus defined as follows:

$$\mathcal{A} = \Big\{ \underbrace{\bar{\boldsymbol{a}} + \mathrm{Diag}(\boldsymbol{\delta})\boldsymbol{\theta}}_{\boldsymbol{a}(\boldsymbol{\theta})} \ \Big| \quad \boldsymbol{\theta} \in \Theta := \{\boldsymbol{\theta} \in \mathbb{R}^n : \|\boldsymbol{\theta}\|_\infty \leq 1, \quad \|\boldsymbol{\theta}\|_1 \leq \Gamma\} \Big\}.$$

We see that the uncertainty set Θ is a polytope. If we further restrict ourselves to the case of nonnegative deviations (it is often possible to reduce to this case, the general case with unsigned deviations is left to the reader), the budgeted uncertainty model can be rewritten to match the conic inequality form (5):

$$\mathcal{A} = \{\boldsymbol{a}(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \Theta\}, \quad \text{where } \Theta := \{\boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{0} \leq \boldsymbol{\theta} \leq \boldsymbol{1}, \quad \boldsymbol{1}^T \boldsymbol{\theta} \leq \Gamma\}. \tag{6}$$

The budgeted uncertainty set (6) was first studied by Bertsimas and Sim, who gave a combinatorial construction of $\mathcal{A}$: it is easy to check that $\mathcal{A} = \mathbf{conv}\,\hat{\mathcal{A}}$, where

$$\hat{\mathcal{A}} = \{\boldsymbol{a} \in \mathbb{R}^n | \ \ a_i \in \{\bar{a}_i, \bar{a}_i + \delta_i\}, \ \forall i \in [n]; \quad |\{i : a_i = \bar{a}_i + \delta_i\}| \leq \Gamma\}.$$

In words, the model (6) protects against all scenarios $\boldsymbol{\theta}$ where each element of $\boldsymbol{a}(\boldsymbol{\theta})$ is equal to either the nominal value a_i or the perturbed value $a_i + \delta_i$, and the number of perturbed values is $\leq \Gamma$.

Example:
Robust knapsack with uncertainty on the weights. Consider the following (continuous) nominal knapsack problem:

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n_+}{\textbf{maximize}} \quad & \boldsymbol{p}^T\boldsymbol{x} \\ s.t. \quad & \boldsymbol{w}^T\boldsymbol{x} \leq W, \end{aligned}$$

where w_i is the weight of one unit of item $i \in [n]$, W is the capacity of the knapsack, and p_i is the profit earned per unit of item i in the knapsack. The goal is to select the quantity $x_i \geq 0$ of each item to be packed in the knapsack, so as to maximize the profit.
We are going to construct the robust counterpart of this problem, when we assume that the vector of weights lies in a budgeted uncertainty set $\mathcal{W}$. This means that some of the items may actually weigh more than in the nominal scenario, but we look for a packing which is feasible for all $\boldsymbol{w} \in \mathcal{W}$, where

$$\mathcal{W} = \Big\{ \bar{\boldsymbol{w}} + \mathrm{Diag}(\boldsymbol{\delta})\boldsymbol{\theta} \mid \quad \boldsymbol{0} \leq \boldsymbol{\theta} \leq \boldsymbol{1}, \quad \boldsymbol{1}^T\boldsymbol{\theta} \leq \Gamma \Big\}.$$ #1

We can apply the result of Theorem 2, by setting $P = \mathrm{Diag}(\delta)$, $F = [-I, I, \boldsymbol{1}]^T$, and $\boldsymbol{h} = [\boldsymbol{0}^T, \boldsymbol{1}^T, \Gamma]^T$:

$$\big(\boldsymbol{w}^T\boldsymbol{x} \leq W, \quad \forall \boldsymbol{w} \in \mathcal{W}\big) \iff \exists (\boldsymbol{y}, \boldsymbol{z}, \zeta) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}_+ : \begin{cases} \bar{\boldsymbol{w}}^T\boldsymbol{x} + \boldsymbol{1}^T\boldsymbol{z} + \Gamma\zeta \leq W \\ \boldsymbol{z} - \boldsymbol{y} + \zeta\boldsymbol{1} = \mathrm{Diag}(\delta)\boldsymbol{x} \end{cases}$$

Finally, we observe that $\boldsymbol{y}$ plays the role of a slack variable, and we can rewrite the robust knapsack problem as

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{z},\zeta}{\textbf{maximize}} \quad & \boldsymbol{p}^T\boldsymbol{x} \\ s.t. \quad & \bar{\boldsymbol{w}}^T\boldsymbol{x} + \boldsymbol{1}^T\boldsymbol{z} + \Gamma\zeta \leq W \\ & z_i + \zeta \geq \delta_i x_i, \quad \forall i \in [n] \\ & \boldsymbol{x} \geq \boldsymbol{0}, \ \boldsymbol{z} \geq \boldsymbol{0}, \zeta \geq 0. \end{aligned}$$

3.2.2 Ellipsoidal Uncertainty

Another commonly used uncertainty set is the ellipsoidal uncertainty set:

$$\mathcal{A} = \Big\{ \underbrace{\bar{\boldsymbol{a}} + P\boldsymbol{\theta}}_{\boldsymbol{a}(\boldsymbol{\theta})} \mid \boldsymbol{\theta} \in \Theta \Big\} \quad \text{where} \quad \Theta := \{\boldsymbol{\theta} \in \mathbb{R}^k : \|\boldsymbol{\theta}\|_2 \leq 1\} = \Big\{ \boldsymbol{\theta} \in \mathbb{R}^n : \begin{bmatrix} -I \\ \boldsymbol{0}^T \end{bmatrix} \boldsymbol{\theta} \preceq_{\mathbb{L}^{n+1}_+} \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix} \Big\},$$

which also matches the conic inequality form (5).

The ellipsoidal model naturally arises when we assume that $\boldsymbol{a}(\boldsymbol{\theta})$ is an affine function of $\boldsymbol{\theta}$, i.e., $\boldsymbol{a}(\boldsymbol{\theta}) = \bar{\boldsymbol{a}} + B\boldsymbol{\theta}$, and $\boldsymbol{\theta} \sim \mathcal{N}(0, \Sigma)$ is a random variable drawn from a multivariate normal distribution with zero mean and variance-covariance matrix $\Sigma \succeq 0$. Then, it can be seen that $\boldsymbol{\theta}$ is equal (in distribution) to $H\boldsymbol{\theta}'$, where $\Sigma = HH^T$, and $\boldsymbol{\theta}' \sim \mathcal{N}(0, I)$ is a standard Gaussian vector (with i.i.d. entries). It is a folklore result from statistics that $\boldsymbol{\theta}'$ lies with probability $1-\alpha$ in the ball $\{\boldsymbol{\theta}' \in \mathbb{R}^k : \|\boldsymbol{\theta}'\|_2 \leq \kappa\}$, where κ is a parameter that depends only[1] on the confidence level $\alpha \in (0,1)$. If we choose as uncertainty set the set of all scenarios in this $(1-\alpha)-$confidence region, we obtain the uncertainty model

$$\mathcal{A} = \{\bar{\boldsymbol{a}} + B\boldsymbol{\theta} \mid \quad \boldsymbol{\theta} = H\boldsymbol{\theta}', \ \|\boldsymbol{\theta}'\|_2 \leq \kappa\} = \Big\{ \bar{\boldsymbol{a}} + \frac{BH}{\kappa}\boldsymbol{\theta}'' \ \Big| \quad \|\boldsymbol{\theta}''\|_2 \leq 1 \Big\}.$$

[1] κ corresponds to the $(1-\alpha)$th percentile of the χ-distribution with k degrees of freedom.

4 Robust Counterpart of SOCP

In this section, we show that it is possible to reformulate the semi-infinite SOC constraint

$$\|A(\boldsymbol{\theta})\boldsymbol{x} + \boldsymbol{b}(\boldsymbol{\theta})\|_2 \leq t, \quad \forall \boldsymbol{\theta} \in \Theta := \{\boldsymbol{\theta} \in \mathbb{R}^k : \|\boldsymbol{\theta}\|_2 \leq 1\} \tag{7}$$

as an LMI. Here, we assume that $A(\cdot)$ and $\boldsymbol{b}(\cdot)$ are affine functions of the uncertain parameter $\boldsymbol{\theta}$. Also, note that there is no uncertainty in the RHS of the constraint. There are many other tractable cases of robust SOCPs and robust SDPs, and we refer the reader to [1] for an exhausitve description of tractable cases.

Our assumption that $A(\boldsymbol{\theta})$ and $\boldsymbol{b}(\boldsymbol{\theta})$ are affine w.r.t. $\boldsymbol{\theta}$ means that there exists some matrices $A_0, \ldots, A_k \in \mathbb{R}^{m\times n}$ and some vectors $\boldsymbol{b}_0, \ldots, \boldsymbol{b}_k \in \mathbb{R}^m$ such that

$$A(\boldsymbol{\theta}) = A_0 + \sum_j \theta_j A_j, \qquad \boldsymbol{b}(\boldsymbol{\theta}) = \boldsymbol{b}_0 + \sum_j \theta_j \boldsymbol{b}_j.$$

Hence, $A(\boldsymbol{\theta})\boldsymbol{x} + \boldsymbol{b}(\boldsymbol{\theta}) = A_0\boldsymbol{x} + \boldsymbol{b}_0 + \sum_j \theta_j(A_j\boldsymbol{x} + \boldsymbol{b}_j) = \boldsymbol{y}_0(\boldsymbol{x}) + L(\boldsymbol{x})\boldsymbol{\theta}$, where

$$\boldsymbol{y}_0(\boldsymbol{x}) = A_0\boldsymbol{x} + \boldsymbol{b}_0 \in \mathbb{R}^m \qquad \text{and} \quad L(\boldsymbol{x}) = [A_1\boldsymbol{x} + \boldsymbol{b}_1, \ldots, A_k\boldsymbol{x} + \boldsymbol{b}_k] \in \mathbb{R}^{m\times k}$$

are affine functions of $\boldsymbol{x}$. So, the semi-infinite constraint (7) can be rewritten as

$$\|\boldsymbol{y}_0(\boldsymbol{x}) + L(\boldsymbol{x})\boldsymbol{\theta}\| \leq t, \quad \forall \|\boldsymbol{\theta}\| \leq 1. \tag{8}$$

To reformulate this constraint as an LMI, we will need two lemmas:

Lemma 3.

$$\inf_{\|\boldsymbol{\theta}\|\leq 1} s\boldsymbol{v}^T L\boldsymbol{\theta} = \inf_{\|\boldsymbol{z}\|\leq s} \boldsymbol{v}^T L\boldsymbol{z}$$

Proof. By using the Cauchy-Schwarz inequality, it is easy to see that the optimal values of both problems are the same, and equal to $-\|s\, L^T\boldsymbol{v}\|$. □

The second lemma is a fundamental result, which gives conditions under which a quadratic inequality is a consequence of another quadratic inequality. It has many applications, in particular in control theory, and will be proved in the exercises.

Theorem 4 (S-lemma)**.** *Let A, B be symmetric matrices, and assume that $\exists \boldsymbol{x}_0 : \boldsymbol{x}_0^T A\boldsymbol{x}_0 > 0$. Then,*

$$\Big(\boldsymbol{x}^T A\boldsymbol{x} \geq 0 \implies \boldsymbol{x}^T B\boldsymbol{x} \geq 0\Big) \iff \Big(\exists \lambda \geq 0 : B - \lambda A \succeq 0\Big). \tag{9}$$

By using a Schur complement, we can reformulate the semi-infinite constraint (8) as:

$$\begin{bmatrix} t & (\boldsymbol{y}_0(\boldsymbol{x}) + L(\boldsymbol{x})\boldsymbol{\theta})^T \\ \boldsymbol{y}_0(\boldsymbol{x}) + L(\boldsymbol{x})\boldsymbol{\theta} & t\, I \end{bmatrix} \succeq 0, \quad \forall \|\boldsymbol{\theta}\| \leq 1.$$

Now, we use the definition of positive semidefinite matrices (the associated quadratic form is nonnegative everywhere). If we evaluate the quadratic form associated with the above matrix at $(s, \boldsymbol{v}) \in \mathbb{R}^{m+1}$, we get:

$$s^2 t + 2s(\boldsymbol{y}_0(\boldsymbol{x}) + L(\boldsymbol{x})\boldsymbol{\theta})^T\boldsymbol{v} + t\|\boldsymbol{v}\|^2 \geq 0, \quad \forall \|\boldsymbol{\theta}\| \leq 1, \forall s \in \mathbb{R}, \forall \boldsymbol{v} \in \mathbb{R}^m.$$

By Lemma 3, we can replace the minimization over $\|\boldsymbol{\theta}\| \leq 1$ by a minimization over $\|\boldsymbol{z}\| \leq s$:

$$s^2 t + 2s\boldsymbol{y}_0(\boldsymbol{x})^T\boldsymbol{v} + 2\boldsymbol{v}^T L(\boldsymbol{x})\boldsymbol{z} + t\|\boldsymbol{v}\|^2 \geq 0, \quad \forall s \in \mathbb{R}, \forall \boldsymbol{v} \in \mathbb{R}^m, \forall \|\boldsymbol{z}\| \leq s.$$

Now, we observe that the above constraint means that a quadratic inequality implies another quadratic inequality:

$$\begin{bmatrix} s \\ \boldsymbol{v} \\ \boldsymbol{z} \end{bmatrix}^T \begin{pmatrix} 1 & & \\ & 0 & \\ & & -I \end{pmatrix} \begin{bmatrix} s \\ \boldsymbol{v} \\ \boldsymbol{z} \end{bmatrix} \geq 0 \quad \Longrightarrow \quad \begin{bmatrix} s \\ \boldsymbol{v} \\ \boldsymbol{z} \end{bmatrix}^T \begin{pmatrix} t & \boldsymbol{y}_0(\boldsymbol{x})^T & \\ \boldsymbol{y}_0(\boldsymbol{x}) & t\,I & L(\boldsymbol{x}) \\ & L(\boldsymbol{x})^T & \end{pmatrix} \begin{bmatrix} s \\ \boldsymbol{v} \\ \boldsymbol{z} \end{bmatrix} \geq 0.$$

This is the required form to use the $S-$lemma. Finally, we have proved:

Theorem 5. *The semi-infinite SOC constraint* (7) *is equivalent to the following linear matrix inequality:*

$$\Big(\|\boldsymbol{y}_0(\boldsymbol{x}) + L(\boldsymbol{x})\boldsymbol{\theta}\| \leq t, \quad \forall \|\boldsymbol{\theta}\| \leq 1 \Big) \iff \exists \lambda \geq 0 : \begin{pmatrix} t-\lambda & \boldsymbol{y}_0(\boldsymbol{x})^T & 0 \\ \boldsymbol{y}_0(\boldsymbol{x}) & t\,I & L(\boldsymbol{x}) \\ 0 & L(\boldsymbol{x})^T & \lambda I \end{pmatrix} \succeq 0.$$

5 Adjustable Robust Counterpart of two-stage problems

In many applications, the decision maker has the possibility to react to the uncertainty, by adjusting her decision after the uncertainty is revealed. Let the nominal problem be an LP of the form

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{y}}{\textbf{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{y} && \text{(10a)} \\ s.t. \quad & A\boldsymbol{x} \leq \boldsymbol{b} && \text{(10b)} \\ & T\boldsymbol{x} + W\boldsymbol{y} \leq \boldsymbol{h}, && \text{(10c)} \end{aligned}$$

where $\boldsymbol{x}$ are the first stage variables, which represent *here-and-now* decisions, while $\boldsymbol{y}$ are the second stage variables, which represent further decisions that can be *adjusted* after the uncertainty is revealed. The constraint (10b) defines the polytope $\mathcal{X}$ such that $\boldsymbol{x} \in \mathcal{X}$, while the constraint (10c) defines a coupling between $\boldsymbol{x}$ and $\boldsymbol{y}$, and gives the set $\mathcal{Y}(\boldsymbol{x})$ of allowed second-stage decisions $\boldsymbol{y}$, as a function of $\boldsymbol{x}$.

Now, we assume that the data T, and $\boldsymbol{h}$ is uncertain (note: we assume a fixed recourse matrix, that is, W is known with certainty). Then, the robust counterpart of (10a) is defined as:

$$\begin{aligned} \underset{\boldsymbol{x}}{\textbf{minimize}} \quad & \sup_{\boldsymbol{\theta}\in\Theta} \inf_{\boldsymbol{y}} \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{y} && \text{(11)} \\ s.t. \quad & A\boldsymbol{x} \leq \boldsymbol{b} \\ & T(\boldsymbol{\theta})\boldsymbol{x} + W\boldsymbol{y} \leq \boldsymbol{h}(\boldsymbol{\theta}), \ \forall \boldsymbol{\theta} \in \Theta \end{aligned}$$

and can be interpreted as a three-stage game: first, the decision maker selects the vector $\boldsymbol{x} \in \mathcal{X}$. Then, an adversary picks a scenario $\boldsymbol{\theta} \in \Theta$, and finally the decision maker observes $\boldsymbol{\theta}$ and can adjust her strategy, by selecting the vector $\boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x},\boldsymbol{\theta}) := \{\boldsymbol{y} |\ W\boldsymbol{y} \leq \boldsymbol{h}(\boldsymbol{\theta}) - T(\boldsymbol{\theta})\boldsymbol{x}\}$.

Since the second stage decisions depend on $\boldsymbol{\theta}$, it can be seen that this problem is actually equivalent to the following problem:

$$\begin{aligned} \underset{\boldsymbol{x},\ \boldsymbol{y}(\boldsymbol{\theta})}{\textbf{minimize}} \quad & \sup_{\boldsymbol{\theta}\in\Theta}\ \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{y}(\boldsymbol{\theta}) \\ s.t. \quad & A\boldsymbol{x} \leq \boldsymbol{b} \\ & T(\boldsymbol{\theta})\boldsymbol{x} + W\boldsymbol{y}(\boldsymbol{\theta}) \leq \boldsymbol{h}(\boldsymbol{\theta}),\ \forall \boldsymbol{\theta}\in\Theta, \end{aligned} \tag{12}$$

where the decision variable $\boldsymbol{y}(\boldsymbol{\theta})$ is now *a function* of $\boldsymbol{\theta} \in \Theta$! Optimizing over a set of functions is a very difficult task, and the above adjustable robust counterpart is unfortunately intractable. However, we get an approximation to this problem when we restrict our attention to *affine decision rules* of the form $\boldsymbol{y}(\boldsymbol{\theta}) = \boldsymbol{y}_0 + Y\boldsymbol{\theta}$. This yields a simpler, one-stage robust optimization problem in the variables $\boldsymbol{x}, \boldsymbol{y}_0$ and Y:

$$\begin{aligned} \underset{\boldsymbol{x},\ \boldsymbol{y}_0,\ Y}{\textbf{minimize}} \quad & \sup_{\boldsymbol{\theta}\in\Theta}\ \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{p}^T(\boldsymbol{y}_0 + Y\boldsymbol{\theta}) \\ s.t. \quad & A\boldsymbol{x} \leq \boldsymbol{b} \\ & T(\boldsymbol{\theta})\boldsymbol{x} + W(\boldsymbol{y}_0 + Y\boldsymbol{\theta}) \leq \boldsymbol{h}(\boldsymbol{\theta}),\ \forall \boldsymbol{\theta}\in\Theta, \end{aligned} \tag{13}$$

The first-stage solution $\boldsymbol{x}^*$ of the above *affinely adjustable robust counterpart* is a conservative solution to the intractable adjustable robust counterpart (11): It can be used as a safe approximation of the solution of (11), where *safe* means that we know that for all realization $\boldsymbol{\theta} \in \Theta$ of the uncertainty, there exists an adjustment $\boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}^*, \boldsymbol{\theta})$. Moreover, $\boldsymbol{x}^*$ minimizes an upper bound of Problem (11), because restricting to affine decision rules can be interpreted as a relaxation of Problem (12). This approach has been implemented succesfully in many applications, in particular in inventory management.

We claim that if $T(\boldsymbol{\theta})$ and $\boldsymbol{h}(\boldsymbol{\theta})$ are affine functions of $\boldsymbol{\theta}$, then the above problem has the same form as the robust counterpart of the LP studied in Section 2. To see this, look at the ith row of the semi-infinite coupling constraint:

$$\boldsymbol{t}_i(\boldsymbol{\theta})^T\boldsymbol{x} + \boldsymbol{w}_i^T\boldsymbol{y}_0 + \boldsymbol{w}_i^T Y\boldsymbol{\theta} \leq h_i(\theta),\ \forall \boldsymbol{\theta}\in\Theta, \tag{14}$$

where $\boldsymbol{t}_i(\boldsymbol{\theta})^T$, $\boldsymbol{w}_i^T$, and $h_i(\boldsymbol{\theta})$ are the ith rows of $T(\boldsymbol{\theta})$, W, and $\boldsymbol{h}(\boldsymbol{\theta})$, respectively. We recognize a linear semi-infinite constraint of the same form as the one studied in Section 3.1 (albeit the dependency w.r.t. $\boldsymbol{\theta}$ in the RHS). So, for example, if $\Theta = \{\boldsymbol{\theta} \mid F\boldsymbol{\theta} \preceq_K \boldsymbol{f}\}$ we could use Theorem 2 to reformulate (14) as a compact system of conic inequalities. We find it simpler to directly derive the counterpart by mimicking the proof of Theorem 2, which avoids a tedious vectorization of the matrix variable Y. Assume that $\boldsymbol{t}_i(\boldsymbol{\theta}) = \boldsymbol{t}_i + T_i\boldsymbol{\theta}$ and $h_i(\boldsymbol{\theta}) = \eta_i + \boldsymbol{g}_i^T\boldsymbol{\theta}$. Then, constraint (14) can be rewritten as:

$$\sup_{\{\boldsymbol{\theta} \mid F\boldsymbol{\theta}\preceq_K \boldsymbol{f}\}} \boldsymbol{\theta}^T(T_i^T\boldsymbol{x} + Y^T\boldsymbol{w}_i - \boldsymbol{g}_i) \leq \eta_i - \boldsymbol{t}_i^T\boldsymbol{x} - \boldsymbol{w}_i^T\boldsymbol{y}_0$$

Then, provided strong duality holds, we get the equivalent constraint

$$\begin{aligned} \inf_{\boldsymbol{z}_i} \quad & \boldsymbol{z}_i^T\boldsymbol{f} \leq \eta_i - \boldsymbol{t}_i^T\boldsymbol{x} - \boldsymbol{w}_i^T\boldsymbol{y}_0 \\ s.t. \quad & F^T\boldsymbol{z}_i = T_i^T\boldsymbol{x} + Y^T\boldsymbol{w}_i - \boldsymbol{g}_i \\ & \boldsymbol{z}_i \succeq_{K^*} \boldsymbol{0}. \end{aligned}$$

It remains to handle the supremum w.r.t. $\boldsymbol{\theta}$ in the objective function of (13). This can be handled by using the same reasoning as above:

$$\left(\sup_{\boldsymbol{\theta}\in\Theta} \boldsymbol{p}^T Y\boldsymbol{\theta}\ \leq t\right) \iff \exists \boldsymbol{z}_0 \succeq_{K^*} \boldsymbol{0} : \left\{ \begin{array}{l} \boldsymbol{z}_0^T\boldsymbol{f} \leq t \\ F^T\boldsymbol{z}_0 = Y^T\boldsymbol{p}. \end{array}\right.$$

Finally, we obtain the following result:

Theorem 6. *Let* $\Theta = \{\boldsymbol{\theta} \mid F\boldsymbol{\theta} \preceq_K \boldsymbol{f}\}$*, and assume that the ith row of* $T(\boldsymbol{\theta})$ *and the ith component of* $\boldsymbol{h}(\boldsymbol{\theta})$ *are affine functions, of the form* $\boldsymbol{t}_i(\boldsymbol{\theta})^T = (\boldsymbol{t}_i + T_i\boldsymbol{\theta})^T$*, and* $h_i(\boldsymbol{\theta}) = \eta_i + \boldsymbol{g}_i^T\boldsymbol{\theta}$*, for all* $i \in [m]$*. Assume further that the conic inequality* $F\boldsymbol{\theta} \preceq_K \boldsymbol{f}$ *is essentially strictly feasible.*

Then, a safe approximation $\boldsymbol{x}^*$ *of Problem* (11)*, that solves the affinely adjustable robust counterpart* (13)*, can be computed by solving the following conic optimization problem:*

$$\begin{aligned}
\underset{\substack{\boldsymbol{x},\, \boldsymbol{y}_0,\, Y\\ \boldsymbol{z}_0,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_m}}{\textbf{minimize}} \quad & \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{p}^T\boldsymbol{y}_0 + \boldsymbol{z}_0^T\boldsymbol{f} \\
s.t. \quad & A\boldsymbol{x} \leq \boldsymbol{b} \\
& \boldsymbol{t}_i^T\boldsymbol{x} + \boldsymbol{w}_i^T\boldsymbol{y}_0 + \boldsymbol{z}_i^T\boldsymbol{f} \leq \eta_i, \quad \forall i \in [m] \\
& F^T\boldsymbol{z}_i = T_i^T\boldsymbol{x} + Y^T\boldsymbol{w}_i - \boldsymbol{g}_i, \quad \forall i \in [m] \\
& F^T\boldsymbol{z}_0 = Y^T\boldsymbol{p} \\
& \boldsymbol{z}_0 \succeq_{K^*} \boldsymbol{0},\ \boldsymbol{z}_1 \succeq_{K^*} \boldsymbol{0}, \ldots,\ \boldsymbol{z}_m \succeq_{K^*} \boldsymbol{0}.
\end{aligned}$$

References

[1] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization.* Princeton University Press, 2009.