Robust optimization methods for chance constrained, simulation-based, and bilevel problems

Yanikoglu, I.

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Robust Optimization Methods for Chance Constrained, Simulation-Based, and Bilevel Problems

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Ruth First zaal van de Universiteit op dinsdag 2 september 2014 om 14.15 uur door

İhsan Yanikoğlu

geboren op 4 augustus 1985 te İzmir, Turkije

Promotiecommissie:

Promotor:	prof.dr.ir. Dick den Hertog
Copromotor:	dr. Daniel Kuhn
Overige leden:	prof.dr. Aharon Ben-Tal prof.dr. Etienne de Klerk prof.dr. Goos Kant dr. Ruud Brekelmans dr. Juan Vera Lizcano

Robust Optimization Methods for Chance Constrained, Simulation-Based, and Bilevel Problems

Acknowledgements

I would like to express my most sincere thanks to my supervisor Dick den Hertog. He is one of the most influential people I have met in my entire life, who truly changed it in many different ways. I would like to thank him once more for everything he has made at all stages of my life in Tilburg.

I would like to thank Daniel Kuhn for giving me the opportunity to work with him in the final year of my Ph.D. and accepting to co-supervise my thesis. I also express my gratitude to him and his colleagues in EPFL for the friendly environment they have created during my three months stay in Lausanne.

I thank to my co-authors Jack Kleijnen and Bram Gorissen for their valuable comments and discussions during the realization of the two chapters in this thesis.

A special mention goes to Gül Gürkan who approved my application to Tilburg University in 2009, and arranged everything in the first two years.

I would like to thank CentER Research Institute for their financial support throughout my studies. CentER made everything easier for me in the past five years.

I thank committee members Aharon Ben-Tal, Etienne de Klerk, Goos Kant, Ruud Brekelmans, and Juan Vera Lizcano for accepting to read and review this thesis.

Finally, I would like to thank to my family for their everlasting support and love.

İhsan Yanıkoğlu
Tilburg, June 2014

Contents

CHAPTER 1

Introduction

Robust optimization (RO) is a young and active research field that has been mainly developed in the course of last fifteen years. For a quick overview of the associated literature, we refer to the survey papers by Ben-Tal and Nemirovski (2002); Bertsimas et al. (2011); Beyer and Sendhoff (2007). Additional details on the history of RO that is in line with the research of this thesis shall be discussed in later chapters.

RO is well-known because it yields computationally tractable methods for uncertain optimization problems. Unlike its counterparts, i.e., dynamic and stochastic programming, it does not suffer from the curse of dimensionality. However, despite all its computational advantages, it is remarkable that practical use of RO in real-life is still lagging behind. This is mainly because of some issues and shortcomings that hinder successful applications of RO in real-life. In this thesis, our goal is to overcome some of these shortcomings and to help practitioners for successfully applying RO in practice. In the next chapter, several practical topics are discussed in separate sections. In addition, a practical guide, that shows a roadmap on how to apply RO in practice, is given. In the remaining chapters of the thesis, we propose and analyze tractable RO methodologies for several classes of optimization problems.

In this chapter, we give a brief introduction on optimization under uncertainty, as well as the contributions of our research and the overview of the following chapters. Section 1.1 describes the sources of uncertainties in mathematical optimization problems and the optimization techniques to deal with these uncertainties. Section 1.2 gives a brief introduction on RO. Section 1.3 describes the concept of adjustable RO. Section 1.4 discusses the contribution of this thesis.

1.1 Optimization Under Uncertainty

Mathematical optimization problems often have uncertainty in problem parameters because of measurement/rounding, estimation/forecasting, or implementation errors. We describe these errors in detail below.

Measurement/rounding errors are often caused when an actual measurement is rounded to a nearest value according to a rule, e.g., the nearest tenth or hundredth, or when the actual value of the parameter cannot be measured with a high precision as it appears in reality. For example, if the reported parameter value is 9.5 according to the nearest tenth, then the actual value can be anywhere between 9.45 and 9.55, i.e., it is uncertain.

Estimation/forecasting errors come from the lack of true knowledge about the problem parameter or the impossibility to estimate the true characteristics of the actual data. For example, demand and cost parameters are often subject to such estimation/forecasting errors.

Implementation errors are often caused by "ugly" reals that can be hardly implemented with the same precision in reality. For example, suppose the optimal voltage in a circuit, that is calculated by an optimization tool, is 2.782169. The decimal part of this optimal solution can be hardly implemented in practice, since you cannot provide the same precision.

Optimization based on nominal values often lead to "severe" infeasibilities. Notice that a small uncertainty in the problem data can make the nominal solution completely useless. A case study in Ben-Tal and Nemirovski (2000) shows that perturbations as low as 0.01% in problem coefficients result constraint violations more than 50% in 13 out of 90 *NETLIB* Linear Programming problems considered in the study. In 6 of this 13 problems violations were over 100%, where 210,000% being the highest (i.e., seven scale higher than the tested uncertainty). Therefore, a practical optimization methodology that proposes immunity against uncertainty is needed when the uncertainty heavily affects the quality of the nominal solution.

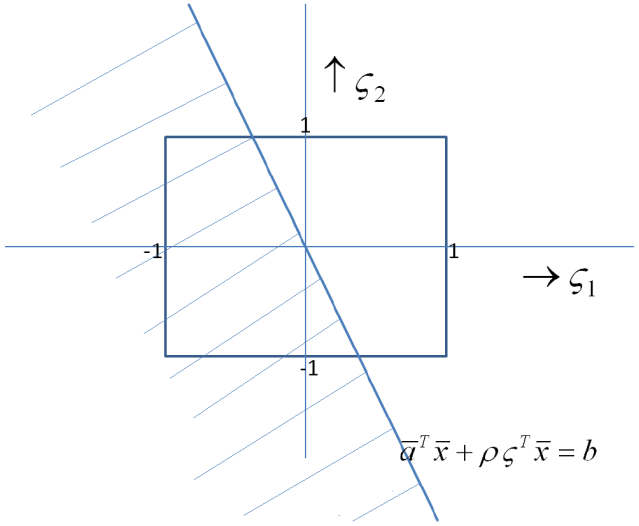
Illustrative example on flaw of using nominal values. Consider an uncertain linear optimization problem with a single constraint:

$$a^\top x \leq b, \tag{1.1}$$

where $a = \bar{a} + \rho\zeta$ is the vector of uncertain coefficients and $\bar{a}$ being the nominal

vector, $\zeta \in \mathbb{R}^2$ is the uncertain parameter that is uniformly distributed in a unit box $||\zeta||_\infty \leq 1$, and ρ is a scalar shifting parameter. Now (say) we ignore the uncertainty in the constraint coefficients and solve the associated problem according to the nominal data, i.e., $a = \bar{a}$, and assume that the constraint is binding for the associated nominal optimal solution $\bar{x}$, i.e., $\bar{a}\bar{x} = b$. Figure 1.1 shows the original constraint $[a^\top x \leq b]$ in the uncertainty space when x is fixed to the nominal optimal solution $\bar{x}$.

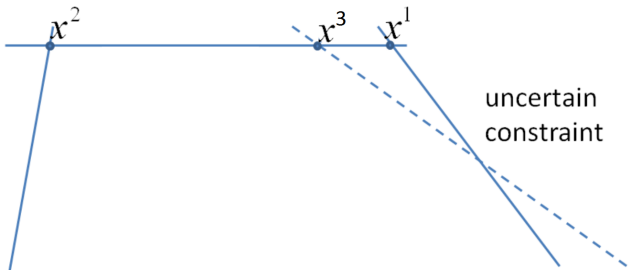
Figure 1.1 – Feasible region of the uncertain constraint $(\bar{a} + \rho\zeta)^\top \bar{x} \leq b$ in the uncertainty space $[-1, 1]^2$.



The solid line in Figure 1.1 represents ζ values where the uncertain constraint is binding when x is fixed to the nominal solution $\bar{x}$, and the dashed lines represent the feasible uncertainty region for the same constraint. Therefore, the area that is determined by the intersection of the unit box with the dashed region gives the subset for which the nominal $\bar{x}$ is robust. From the figure we can conclude that the probability of violating this constraint can be as high as 50%, since ζ follows a uniform distribution. This shows that uncertainty may severely affect the quality of the nominal solution, and there exists a crucial need for an optimization methodology that yields solutions that are immunized against the uncertainty.

Now let us consider the following figure that presents another illustrative example. In Figure 1.2, there are three constraints and their binding values are represented by the solid lines. The constraint on the right-hand side is uncertain, and the other two are certain. For the uncertain constraint, the solid line represents the binding

Figure 1.2 – Effects of uncertainty on feasibility and optimality performance of solutions (this figure is represented in the decision variable space).



value of the constraint for the nominal data and the dashed line represents the same for a different realization of the uncertain data. Notice that, different than in Figure 1.1, in Figure 1.2 we are in the space of the decision variable x. We assume that the problem at hand is an uncertain linear problem, therefore, the optimal solutions are obtained at the extreme points of the feasible region where the constraints are binding. Suppose x^1 denotes the unique nominal optimal solution of the problem. It is easy to see that x^1 may be highly infeasible when the associated constraint is uncertain. The new (robust) optimal solution may become x^3. Now consider the case where x^1 and x^2 are both optimal for the nominal data, i.e., the optimal facet is the line segment that connects x^1 and x^2. In this case, the decision maker would always prefer x^2 over x^1, since its feasibility performance is less affected by the uncertainty. This shows that staying away from "risky" solutions that have uncertain binding constraints may be beneficial.

There are two complementary approaches in optimization that deals with data uncertainty, namely, RO and *stochastic optimization* (SO). Each method has its own assumptions. To begin with, basic SO has the following assumptions (Ben-Tal et al., 2009, p. xiii):

S.1. The underlying probability distribution or a restricted family of distributions of the uncertain parameter must be known.

S.2. The associated family of distributions or the distribution should not change over the considered time horizon that the decisions will be made.

S.3. The decision maker should be ready to accept probabilistic guarantees as the performance measure against the uncertainty.

If these conditions are met and the deterministic counterpart of the stochastic problem is tractable, then SO is the right optimization methodology to solve the problem

at hand. For additional details on SO we refer to Prékopa (1995); Birge and Louveaux (2011); Shapiro and Ruszczyński (2003), and Charnes and Cooper (1959).

On the other hand, the "basic" RO approach has the following three implicit assumptions (Ben-Tal et al., 2009, p. xii):

R.1. All decision variables represent "here and now" decisions: they should get specific numerical values as a result of solving the problem before the actual data "reveals itself".

R.2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set.

R.3. The constraints of the uncertain problem in question are "hard" – the decision maker cannot tolerate violations of constraints when the data is in the prespecified uncertainty set.

It is important to point out that assumption [R.1] can be alleviated by *adjustable robust optimization* (ARO); a brief introduction on ARO will also be given in Section 1.3, but details will follow in Chapters 4 & 5. In addition, assumption [R.3] can be alleviated by *globalized robust optimization* (Ben-Tal et al., 2009, Ch. 3 & 11), as well as by using safe approximations of chance constraints that shall be discussed in Chapter 3.

If we compare the basic versions of RO and SO, the latter seems to be less conservative than the former since it is not worst-case oriented. However, it is important to point out that the SO approach is valid only when the probability distribution is known. Notice that RO does not have such a restriction since it works with uncertainty sets that can be derived by expert opinion or using historical data. Moreover, the RO paradigm is computationally more tractable than the SO approach; for details on such examples we refer to Chapter 3, Ben-Tal et al. (2009, pp. xiii - xv), and Chen et al. (2006). It is important to point out that, although this thesis is based on RO, some chapters are closely related to SO, e.g., Chapter 3 deals with safe approximations of ambiguous chance constraints, and such ambiguous chance constraints can also be analyzed in the SO approach, and Chapter 5 proposes stochastic and robust reformulations of a class of bilevel optimization problems. Chapter 3 shows how SO and RO may substitute each other according to different sets of assumptions on the uncertainty; and Chapter 5 shows how the two approaches may complement

each other under mild assumptions on the uncertainty.

As mentioned before, this thesis is based on the RO paradigm and we devote a separate section on RO below.

1.2 Robust Optimization

> "*A specific and relatively novel methodology for handling mathematical optimization problems with uncertain data*" Ben-Tal et al. (2009)

The objective of RO is to find solutions that are robust to the uncertainty of parameters in a mathematical optimization problem. It requires that the constraints of a given problem should be satisfied for all realizations of the uncertain parameters in a so-called uncertainty set. The robust version of a mathematical optimization problem is generally referred to as the *robust counterpart* (RC) problem. Below we present the RC of an uncertain linear optimization problem:

$$\max_x \{c^\top x : \ Ax \leq b \quad \forall A \in \mathbf{U}\}, \tag{1.2}$$

where $x \in \mathbb{R}^n$ are the decision variables, $c \in \mathbb{R}^n$ are the certain cost coefficients, $b \in \mathbb{R}^n$ is the certain constraint right-hand side, $A \in \mathbb{R}^{m \times n}$ denotes an uncertain matrix that resides in a given uncertainty set $\mathbf{U}$, and the constraints should be satisfied for all the uncertainty in $\mathbf{U}$. In RO we may assume without loss of generality that: 1) the objective is certain; 2) the constraint right-hand side is certain; 3) $\mathbf{U}$ is compact and convex; and 4) the uncertainty is constraint-wise.

Below, we explain the technical reasons of why the above stated four "basic" RO assumptions are not restrictive.

E.1. Suppose the objective coefficients (c) of the optimization problem (1.2) are uncertain and (say) these coefficients reside in the uncertainty set $\mathcal{C}$:

$$\max_x \ \min_{c \in \mathcal{C}} \{c^\top x : Ax \leq b \quad \forall A \in \mathbf{U}\}.$$

Without loss of generality we may assume that the uncertain objective of the optimization problem can be equivalently reformulated as certain:

$$\max_{x \in \mathcal{X}, t} \{t : t - c^\top x \leq 0 \quad \forall c \in \mathcal{C}, \ Ax \leq b \quad \forall A \in \mathbf{U}\},$$

using an epigraphic reformulation and the additional variable $t \in \mathbb{R}$.

E.2. The second assumption is trivial because the uncertain right-hand side of a constraint can always be translated to the left-hand side by introducing an extra variable $x_{n+1} = 1$.

E.3. The uncertainty set $\mathbf{U}$ can be replaced by its convex hull $\text{conv}(\mathbf{U})$ in (1.2), because testing the feasibility of a solution with respect to $\mathbf{U}$ is equivalent to maximizing a linear constraint over $\mathbf{U}$, and it yields the same optimal objective value if the constraint is maximized over $\text{conv}(\mathbf{U})$. For details of the formal proof and the compactness assumption, see (Ben-Tal et al., 2009, pp. 12–13).

E.4. To illustrate that robustness with respect to $\mathbf{U}$ can be formulated constraint-wise, consider a problem with two constraints and with uncertain right-hand sides b_1 and b_2: $x_1 + b_1 \leq 0$, $x_2 + b_2 \leq 0$. Let $\mathbf{U} = \{b \in \mathbb{R}^2 | b_1 \geq 0, b_2 \geq 0, b_1 + b_2 \leq 1\}$ be the uncertainty set, and $U_1 = \{b_1 \in \mathbb{R} | 0 \leq b_1 \leq 1\}$ and $U_2 = \{b_2 \in \mathbb{R} | 0 \leq b_2 \leq 1\}$ are the projections of $\mathbf{U}$ on b_1 and b_2. It is easy to see that robustness of the i-th constraint with respect to $\mathbf{U}$ is equivalent to robustness with respect to U_i, i.e., the uncertainty in the problem data can be modelled constraint-wise. For the general proof, see (Ben-Tal et al., 2009, pp. 11–12).

For uncertain *nonlinear* optimization problems, excluding the third basic assumption [E.3], the other three assumptions [E.1], [E.2], and [E.4] are also without loss of generality.

Computational complexity. Notice that (1.2) has infinitely many constraints with finite number of variables, i.e., it is a semi-infinite optimization problem. Therefore, (1.2) is a computationally challenging problem to be solved in its current version. RO is popular because it proposes computationally tractable reformulations of such semi-infinite optimization problems for many classes of uncertainty sets and problem types (including several classes of nonlinear optimization problems).

There are three important steps to derive the associated tractable RC. We shall explicitly go through these three steps below. For the sake of exposition we shall focus on an uncertain linear optimization problem with a polyhedral uncertainty set, but the associated procedure can be applied for other uncertainty sets and problem types as well.

Three steps to derive tractable RC. As it is mentioned above, the uncertainty

is constraint-wise in RO, and hence we may focus on a single constraint

$$(a + B\zeta)^\top x \leq \beta \quad \forall \zeta \in \mathcal{Z}, \tag{1.3}$$

where $\zeta \in \mathbb{R}^k$ is the "primitive" uncertain parameter residing in the polyhedral uncertainty set $\mathcal{Z} = \{\zeta : D\zeta + d \geq 0\}$. In the left-hand side of (1.3), we use a factor model to formulate the general uncertain parameter a_ζ as an affine function $a + B\zeta$ of the *primitive* uncertainty ζ (i.e., $a_\zeta := a + B\zeta$). To point out, the dimension of the general uncertain parameter a_ζ is often much higher than that of the primitive uncertainty ζ. Notice that (1.3) is equivalent to the following worst-case reformulation:

$$\textbf{Step 1 (Worst-case):}\ a^\top x + \max_{\zeta:\ D\zeta+d\geq 0} (B^\top x)^\top \zeta \leq \beta. \tag{1.4}$$

Then we take the dual of the inner maximization problem in (1.4). Notice that by strong duality the inner maximization problem and its dual yield the same optimal objective value. Therefore, (1.4) is equivalent to

$$\textbf{Step 2 (Duality):}\ a^\top x + \min_y \{d^\top y :\ D^\top y = -B^\top x,\ y \geq 0\} \leq \beta. \tag{1.5}$$

It is important to point out that we can also omit the minimization term in (1.5), since it yields an upper bound for the maximization problem in (1.4). Hence, the final formulation of the RC becomes

$$\textbf{Step 3 (RC):}\ \exists\, y: \quad a^\top x + d^\top y \leq \beta, \quad D^\top y = -B^\top x, \quad y \geq 0. \tag{1.6}$$

Note that the constraints in (1.6) are linear in x and y. A similar procedure can also be applied to derive the RCs for different classes of uncertainty sets and problem types; for additional details on deriving the tractable RCs we refer to Ben-Tal et al. (2009, 2014).

Table 1.1 taken from Ben-Tal et al. (2014) presents the tractable robust counterparts of an uncertain linear optimization problem for different classes of uncertainty sets. These robust counterparts are derived using the three step procedure that is described above, however, we need conic duality instead of LP duality in Step 2 to derive the tractable robust counterparts for the ellipsoidal and the conic uncertainty sets; see the second and the fourth rows of Table 1.1. Similarly, to derive the tractable RC in the fifth row of Table 1.1, we need Fenchel duality in Step 2; see Rockafellar (1997) for details on Fenchel duality, and Ben-Tal et al. (2014) for the formal proof of the associated RC reformulation. Notice that each RC constraint has a positive safe guard in the constraint left-hand side, e.g., $\rho\|B^\top x\|_1$, $\rho\|B^\top x\|_2$, and $d^\top y$; see

Table 1.1 – Tractable Robust Counterparts of $[(a + B\zeta)^\top x \leq \beta \quad \forall \zeta \in \mathcal{Z}]$ for different choices of the uncertainty set $\mathcal{Z}$

Uncertainty region	$\mathcal{Z}$	Robust Counterpart	Tractability
Box	$\|\zeta\|_\infty \leq \rho$	$a^\top x + \rho\|B^\top x\|_1 \leq \beta$	LP
Ball/ellipsoidal	$\|\zeta\|_2 \leq \rho$	$a^\top x + \rho\|B^\top x\|_2 \leq \beta$	CQP
Polyhedral	$D\zeta + d \geq 0$	$\begin{cases} a^\top x + d^\top y \leq \beta \\ D^\top y = -B^\top x \\ y \geq 0 \end{cases}$	LP
Cone (closed, convex, pointed)	$D\zeta + d \in K$	$\begin{cases} a^\top x + d^\top y \leq \beta \\ D^\top y = -B^\top x \\ y \in K^* \end{cases}$	Conic Opt.
Convex functions	$\begin{cases} h_k(\zeta) \leq 0 \\ k = 1, \ldots, k \end{cases}$	$(*)\begin{cases} a^\top x + \sum_k u_k \, h^*(\frac{w^k}{u^k}), \\ \sum_k w^k = B^\top x, \\ u \leq 0 \end{cases}$	Convex Opt.

$(*)$ h^* denotes the convex conjugate function, i.e, $h^*(x) = \sup_y\{x^\top y - h(y)\}$

the tractable RCs in the third column of Table 1.1. These safe guards represent the level of robustness that we introduce to the constraints. We refer to Ben-Tal et al. (2009, pp. 373–388) and Ben-Tal et al. (2014) for the tractable RCs reformulations of several classes of uncertain nonlinear optimization problems. Also see, Gorissen et al. (2014) who propose a method that is based on the optimistic dual reformulation with respect to the full uncertain problem (1.2).

1.3 Adjustable Robust Optimization

As it is stated in Section 1.2, the basic assumption [R.1] (p. 116), can be alleviated by *adjustable robust optimization* (ARO). Different than the "classic" RO that models decision variables as "here and now", in ARO decision variables can also be *"wait and see"* that depends on the data revealing itself before the decision is made. For example, in a multistage uncertain optimization problem, a decision that is made at period t may depend on the data that reveals itself before at time $1, \ldots, t-1$. ARO adopts decision rules to model the associated dependence between the data and the decisions.

Adjustable robust counterpart (ARC). Consider a general uncertain linear optimization problem

$$\min_{x,y} \{c_\zeta^\top x + d^\top y : A_\zeta x + By \leq \beta\} : \zeta \in \mathcal{Z}, \tag{1.7}$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^l$ are the decision vectors; $\zeta \in \mathbb{R}^k$ is the uncertain parameter, $c_\zeta \in \mathbb{R}^n$ and $A_\zeta \in \mathbb{R}^{m\times n}$ are the uncertain coefficients that are affine in ζ; $d \in \mathbb{R}^l$ and $B \in \mathbb{R}^{m\times l}$ are the certain parameters. Notice that we have a fixed recourse formulation in (1.7), i.e., B is a certain coefficient matrix. The RC of (1.7) is the following semi-infinite optimization problem

$$\min_{x,y,t} \{t : \forall \zeta \in \mathcal{Z} : c_\zeta^\top x + d^\top y - t \leq 0,\ A_\zeta x + By \leq \beta\}, \tag{1.8}$$

where $t \in \mathbb{R}$ represents the additional variable that comes from the epigraphic reformulation of the uncertain objective.

As explained earlier, in multi-stage optimization some decision variables can be adjusted at a later moment in time when a portion of the uncertain data ζ reveals itself. For example, in a multi-stage inventory system affected by uncertain demand, the replenishment order of day t is made when we know the actual demands in the preceding days; for practical examples see Ben-Tal et al. (2009, Ch. 14.2.1). Now suppose y in (1.8) denotes such an adjustable variable, i.e., y is a function of ζ as $y(\zeta)$. Therefore, the ARC reformulation is given as follows:

$$\min_{x,y,t} \{t : \forall \zeta \in \mathcal{Z} : c_\zeta^\top x + d^\top y(\zeta) - t \leq 0,\ A_\zeta x + By(\zeta) \leq \beta\}, \tag{ARC}$$

where x is the first-stage decision that is made before ζ is realized, and y denotes the second-stage decision that can be adjusted according to the actual data.

However, ARC is an NP-Hard problem unless we restrict the feasible function space of $y(\zeta)$ to specific classes; see Ben-Tal et al. (2009, Ch. 14) for details. In practice, $y(\zeta)$ is often approximated by affine decision rules:

$$y(\zeta) := y^0 + \sum_{j=1}^{k} y^j \zeta_j, \tag{1.9}$$

because they yield computationally tractable *affinely* ARC (AARC) reformulations (see below), where $y^0 \in \mathbb{R}^l$ denotes the intercept variable, and $y^j \in \mathbb{R}^l$ are the vectors of "information base" variables.

If we suppose that the uncertain coefficient matrix A_ζ is affine in ζ:

$$A_\zeta = A^0 + \sum_{j=1}^{k} \zeta_j A^j, \tag{1.10}$$

then the constraints of the ARC that is adopting (4.31) and (1.10) can be reformulated as

$$A^0 x + By^0 + \sum_{j=1}^{k} (A^j x + By^j)\zeta_j \leq 0 \quad \forall \zeta \in \mathcal{Z}. \tag{AARC}$$

Therefore, the i-th constraint of the ARC is given as follows

$$A_i^0 x + B_i y^0 + a_i^\top \zeta \leq 0 \quad \forall \zeta \in \mathcal{Z}, \quad i \in \{1, \ldots, m\} \tag{1.11}$$

where $a_i^\top = [A_i^1 x + B_i y_i^1,\ A_i^2 x + B_i y_i^2, \ldots,\ A_i^k x + B_i y_i^k]$. Eventually, the tractable reformulation of the ARC constraints can be derived as in Table 1.1, since the resulting formulation (1.11) is affine in x, y, and ζ.

Notice that we adopt affine decision rules in the ARC, but it is important to point out that tractable ARC reformulations for nonlinear decision rules also exist for specific classes; we refer to Ben-Tal et al. (2009, Ch. 14.3) and Georghiou et al. (2010) for such tractable ARC reformulations.

The advantages of ARO can be explained in threefold:

1) ARO is less conservative than the classic RO approach, since it yields more flexible decisions that can be adjusted according to the realized portion of data at a given stage. More precisely, ARO yields optimal objective values that are at least as good as that of the standard RO approach.

2) Aside from introducing additional variables and constraints, *affinely* ARO (AARO) does not introduce additional computational complexity to that of RO, and it can be straightforwardly adopted to the classic RO framework, i.e., AARO is a tractable approach.

3) It has applications in real-life, e.g., supply chain management (Ben-Tal et al., 2005), project management (Ben-Tal et al., 2009, Ex. 14.2.1), and so on.

Adjustable reformulations of integer variables. The linear and nonlinear decision rules proposed by Ben-Tal et al. (2009, Ch. 14) cannot be applied for adjustable

integer variables, since the associated decision rules cannot guarantee to result an integer decision for any given ζ. To the best of our knowledge, there are three alternative approaches to model such variables. Namely, Bertsimas and Caramanis (2010) propose splitting the uncertainty set into subsets, where each subset has its own binary decision; Bertsimas and Georghiou (2013) propose piecewise constant functions to model integer variables; and Hanasusanto et al. (2014) propose approximating the associated mixed integer problems by their corresponding K-adaptability problem in which the decision maker pre-commits to K second-stage policies here-and-now and implements the best of these policies once the uncertain parameters are observed. Similar to Bertsimas and Caramanis, in Chapter 2 we propose some methods to model integer variables via so-called cell-based decision rules, and then in Chapter 4 we present specific applications of the associated decision rules in a class of simulation optimization problems.

1.4 Contribution and Overview

This thesis consists of four self-contained chapters on RO. In this section, we give the contributions of each chapter.

The aim of Chapter 2 is to help practitioners to successfully apply RO in practice. Many practical issues are treated, as: (i) how to choose the uncertainty set? (ii) Should the decision rule be a function of the final or the primitive uncertain parameters? (iii) Should the objective also be optimized for the worst case? (iv) How to deal with integer adjustable variables? (v) How to deal with equality constraints? (vi) What is the right interpretation of "RO optimizes for the worst case"? (vii) How to compare the robustness characteristics of two solutions?

Moreover, we pinpoint several important items that may be helpful for successfully applying RO. Some items are: (i) the robust reformulations of two equivalent deterministic optimization problems may not be equivalent. (ii) Comparing the robust objective value of the robust solution with the nominal objective value of the nominal solution is incorrect when the objective is uncertain. (iii) For several multi-stage problems the normal robust solution, or even the nominal solution, may outperform the adjustable solution both in the worst case and in the average performance when the solution is re-optimized in each stage. We use many small examples to demonstrate the associated practical RO issues and items.

In Chapter 3, we propose a new way to construct uncertainty sets for RO. Our ap-

proach uses the available historical data for the uncertain parameters and is based on goodness-of-fit statistics. It guarantees that the probability that the uncertain constraint holds is at least the prescribed value. Compared to existing safe approximation methods for chance constraints, our approach directly uses the historical data information and leads to tighter uncertainty sets and therefore to better objective values. This improvement is significant especially when the number of uncertain parameters is low. Other advantages of our approach are that it can easily handle joint chance constraints, it can deal with uncertain parameters that are dependent, and it can be extended to nonlinear inequalities. Several numerical examples illustrate the validity of our approach. The limitation of the proposed methodology is that it requires extensive data when the number of uncertain parameters is high.

In Chapter 4, we present a novel combination of RO developed in mathematical programming, and robust parameter design developed in statistical quality control. Robust parameter design uses metamodels estimated from experiments with both controllable and environmental inputs. These experiments may be performed with either real or simulated systems; we focus on simulation experiments. For the environmental inputs, classic robust parameter design assumes known means and covariances, and sometimes even a known distribution. We, however, develop a RO approach that uses only experimental data, so it does not need these classic assumptions. Moreover, we develop 'adjustable' robust parameter design which adjusts the values of some or all of the controllable factors after observing the values of some or all of the environmental inputs. We also propose a decision rule that is suitable for adjustable integer decision variables. We illustrate our novel method through several numerical examples, which demonstrate its effectiveness.

In Chapter 5, we propose approximations of a specific class of robust and stochastic bilevel optimization problems by using primal and dual linear decision rules. The original formulations of these problems are "severely" intractable. The advantages of our approximation method are: i) it is computationally tractable; ii) we do not need to know the underlying probability distribution of uncertain problem parameters; iii) we can estimate the optimality performance of our approximation.

This thesis is based on the following four research papers:

Chapter 2 B. L. Gorissen, İ. Yanıkoğlu and D. den Hertog. Hints for practical robust optimization. *CentER Discussion Paper No. 2013-065*, 2013. (*OMEGA*, Revise and Resubmit)

Chapter 3 İ. Yanıkoğlu, D. den Hertog. Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4), 666–681, 2013.

Chapter 4 İ. Yanıkoğlu, D. den Hertog and J. P. C. Kleijnen. Adjustable robust optimization using metamodels. *CentER Discussion Paper No. 2013-022*, 2013. (Submitted)

Chapter 5 İ. Yanıkoğlu, D. Kuhn. Primal and dual linear decision rules for bilevel optimization problems. Working Paper.

1.5 Disclosure

Each chapter contains ideas and contributions from all its respective authors. In Chapter 2, Sections 2.1, 2.5 and 2.9 are written by me and computational experiments in Section 2.5 are done by me, and the final version of the associated chapter is written by me. In the remainder, except for Section 4.1, Chapters 3, 4, and 5 are all written by me and all computational experiments are conducted by me.

Bibliography

A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.

A. Ben-Tal and A. Nemirovski. Robust optimization – methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.

A. Ben-Tal, B. Golany, A. Nemirovski, and J.-P. Vial. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management*, 7(3):248–271, 2005.

A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, 2009. http://sites.google.com/site/robustoptimization/.

A. Ben-Tal, D. den Hertog, and J.-P. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, Forthcoming, 2014.

D. Bertsimas and C. Caramanis. Finite adaptability in multistage linear optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766, 2010.

D. Bertsimas and A. Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Optimization Online*, Sept, 2013.

D. Bertsimas, D. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

H.-G. Beyer and B. Sendhoff. Robust optimization – a comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33–34):3190–3218, 2007.

J. R. Birge and F. V. Louveaux. *Introduction to Stochastic Programming*. Springer, 2011.

A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.

X. Chen, M. Sim, P. Sun, and J. Zhang. A tractable approximation of stochastic programming via robust optimization. *Operations Research*, 2006.

A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Optimization Online*, 2010.

B. L. Gorissen, A. Ben-Tal, H. Blanc, and D. den Hertog. Technical note - deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Operations Research*, 62(3):672–679, 2014.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. Two-stage robust integer programming. Technical report, 2014.

A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.

R. T. Rockafellar. *Convex analysis*, volume 28. Princeton University Press, 1997.

A. Shapiro and A. P. Ruszczyński. *Stochastic Programming*. Elsevier, 2003.

CHAPTER 2

Hints for Practical Robust Optimization

2.1 Introduction

Real-life optimization problems often contain uncertain data. The reasons for data uncertainty could be measurement/estimation errors that come from the lack of knowledge of the parameters of the mathematical model (e.g., the uncertain demand in an inventory model) or could be implementation errors that come from the physical impossibility to exactly implement a computed solution in a real-life setting. There are two complementary approaches to deal with data uncertainty in optimization, namely robust and stochastic optimization. *Stochastic optimization* (SO) has an important assumption, i.e., the true probability distribution of uncertain data has to be known or estimated. If this condition is met and the deterministic counterpart of the uncertain optimization problem is computationally tractable, then SO is the methodology to solve the uncertain optimization problem at hand. For details on SO, we refer to Prekopa (1995); Birge and Louveaux (2011); Shapiro and Ruszczyński (2003), but the list of references can be easily extended. *Robust optimization* (RO), on the other hand, does not assume that probability distributions are known, but instead it assumes that the uncertain data resides in a so-called uncertainty set. Additionally, basic versions of RO assume "hard" constraints, i.e., constraint violation cannot be allowed for any realization of the data in the uncertainty set. RO is popular because of its computational tractability for many classes of uncertainty sets and problem types. For a detailed overview of the RO framework, we refer to Ben-Tal et al. (2009); Ben-Tal and Nemirovski (2008), and Bertsimas et al. (2011).

RO is a relatively young and active research field, and has been mainly developed in the last 15 years. Especially in the most recent 5 years there have been many publications that show the value of RO in applications in many fields including finance (Lobo, 2000), management science (Ben-Tal and Nemirovski, 1998), supply chain (Bertsimas and Thiele, 2004), healthcare (Fredriksson et al., 2011), engineering (Ben-Tal and Nemirovski, 2002), etc. Indeed, the RO concepts and techniques

are very useful for practice and not difficult to understand for practitioners. It is therefore remarkable that real-life applications are still lagging behind; there is much more potential for real-life applications than has been exploited hitherto. In this chapter we pinpoint several items that are important when applying RO and that are often not well understood or incorrectly applied by practitioners.

The aim of this chapter is to help practitioners to successfully apply RO in practice. Many practical issues are treated, as: (i) how to choose the uncertainty set? (ii) Should the decision rule be a function of the final or the primitive uncertain parameters? (iii) Should the objective also be optimized for the worst case? (iv) How to deal with integer adjustable variables? (v) How to deal with equality constraints? (vi) What is the right interpretation of "RO optimizes for the worst case"? (vii) How to compare the robustness characteristics of two solutions?

We also discuss several important insights and their consequences in applying RO. Examples are: (i) sometimes an uncertainty set is constructed such that it contains the true parameter with a prescribed probability. However, the actual constraint satisfaction probability is generally much larger than the prescribed value, since the constraint also holds for other uncertain parameters that are outside the uncertainty set. (ii) The robust reformulations of two equivalent deterministic optimization problems may not be equivalent. (iii) Comparing the robust objective value of the robust solution with the nominal objective value of the nominal solution is incorrect when the objective is uncertain. (iv) For several multi-stage problems the normal robust solution, or even the nominal solution, may outperform the adjustable solution both in the worst case and in the average performance when the solution is re-optimized in each stage.

The remainder of the chapter is organized as follows. Section 2.2 gives a recipe for applying RO. This recipe contains the important items in this chapter. Section 2.3 presents alternative ways of constructing uncertainty sets. Section 2.4 discusses how to model uncertainties in linear (or affine) decision rules. Section 2.5 proposes a RO method to model adjustable integer variables. Section 2.6 shows that binary variables in big-M type constraints are automatically adjustable. Section 2.7 shows that robust counterparts of equivalent deterministic problems are not necessarily equivalent. Section 2.8 presents some ways to deal with equality constraints. Section 2.9 gives insights about maximin and minimax formulations in RO. Section 2.10 shows two tests to quantify the quality of a robust solution. Section 2.11 shows that static RO with folding horizon can take better decisions than linearly adjustable

RO in multi-stage problems. Section 2.12 summarizes our conclusions, and indicates future research topics.

2.2 Recipe for Robust Optimization in Practice

In this section we first give a brief introduction on RO, and then we give a recipe for applying RO in practice. Important items at each step of the recipe and the scopes of other sections that are related to these items are presented in this section.

For the sake of exposition, we use an uncertain linear optimization problem, but we point out that most of our discussions in this chapter can be generalized for other classes of uncertain optimization problems. The "general" formulation of the uncertain linear optimization problem is as follows:

$$\min_{\mathbf{x}} \{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{d}\}_{(\mathbf{c},\mathbf{A},\mathbf{d})\in\mathcal{U}}, \tag{2.1}$$

where $\mathbf{c}$, $\mathbf{A}$ and $\mathbf{d}$ denote the uncertain coefficients, and $\mathcal{U}$ denotes the user specified uncertainty set. The "basic" RO paradigm is based on the following three assumptions (Ben-Tal et al., 2009, p. xii):

1. All decision variables $\mathbf{x}$ represent "here and now" decisions: they should get specific numerical values as a result of solving the problem before the actual data "reveals itself".

2. The decision maker is fully responsible for consequences of the decisions to be made when, and only when, the actual data is within the prespecified uncertainty set $\mathcal{U}$.

3. The constraints of the uncertain problem in question are "hard" - the decision maker cannot tolerate violations of constraints when the data is in the prespecified uncertainty set $\mathcal{U}$.

Without loss of generality, the objective coefficients ($\mathbf{c}$) and the right-hand side values ($\mathbf{d}$) can be assumed certain, as in Chapter 1. Often there is a vector of primitive uncertainties $\boldsymbol{\zeta} \in \mathcal{Z}$ such that the uncertain parameter $\mathbf{A}$ is a linear function of $\boldsymbol{\zeta}$:

$$\mathbf{A}(\boldsymbol{\zeta}) = \mathbf{A^0} + \sum_{\ell=1}^{L} \zeta_\ell \mathbf{A}^\ell,$$

where $\mathbf{A^0}$ is the nominal value matrix, $\mathbf{A}^\ell$ are the shifting matrices, and $\mathcal{Z}$ is the user specified primitive uncertainty set. The robust reformulation of (2.1) that is generally referred to as the *robust counterpart* (RC) problem, is then as follows:

$$\min_{\mathbf{x}}\Big\{\mathbf{c}^\top \mathbf{x} : \quad \mathbf{A}(\boldsymbol{\zeta})\mathbf{x} \leq \mathbf{d} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}\Big\}.$$

A solution $\boldsymbol{x}$ is called robust feasible if it satisfies the uncertain constraints $[\mathbf{A}(\boldsymbol{\zeta})x \leq \mathbf{d}]$ for all realizations of $\boldsymbol{\zeta}$ in the uncertainty set $\mathcal{Z}$.

In multistage optimization, the first assumption of the RO paradigm can be relaxed. For example, the amount a factory will produce next month is not a "here and now" decision, but a "wait and see" decision that will be taken based on the amount sold in the current month. Some decision variables can therefore be adjusted at a later moment in time according to a decision rule, which is a function of (some or all part of) the uncertain data. The *adjustable* RC (ARC) is given as follows:

$$\min_{\mathbf{x}}\{\mathbf{c}^{\top}\mathbf{x}: \quad \mathbf{A}(\boldsymbol{\zeta})\mathbf{x}+\mathbf{B}\mathbf{y}(\boldsymbol{\zeta}) \leq \mathbf{d} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}\},$$

where $\mathbf{B}$ denotes a certain coefficient matrix (i.e., fixed recourse), $\mathbf{x}$ is a vector of non-adjustable variables, and $\mathbf{y}(\boldsymbol{\zeta})$ is a vector of adjustable variables. Linear decision rules are commonly used in practice:

$$\mathbf{y}(\boldsymbol{\zeta}) := \mathbf{y^0} + \sum_{\ell=1}^{L} \mathbf{y}^{\ell}\zeta_{\ell},$$

where $\mathbf{y^0}$ and $\mathbf{y}^{\ell}$ are the coefficients in the decision rule, which are to be optimized. Notice that we assume a fixed recourse situation for tractability. Another factor that affects the computational tractability of ARC is the type of the decision rule, but we shall focus on this issue later in Section 2.4. Now having introduced the general notation in RO and *adjustable* RO (ARO), we can give a recipe for applying RO in practice.

Practical RO Recipe

Step 0: Solve the nominal problem.

Step 1: **a)** Determine the uncertain parameters.
b) Determine the uncertainty set.

Step 2: Check robustness of the nominal solution.
IF the nominal solution is robust "enough" **THEN** stop.

Step 3: **a)** Determine the adjustable variables.
b) Determine the type of decision rules for the adjustable variables.

Step 4: Formulate the robust counterpart.

Step 5: **IF** an exact or approximate tractable reformulation of the (adjustable) robust counterpart can be derived **THEN** solve it.
ELSE use the adversarial approach.

Step 6: Check quality of the robust solution. **IF** the solution is too conservative **THEN** go to **Step 1b**.

In the remainder of this section, we describe the most important items at each step of this algorithm. Several items need a more detailed description, and this is done in Sections 3–11.

Step 0 (*Solve the nominal problem*). First, we solve the problem with no uncertainty, i.e., the nominal problem.

Step 1a (*Determine uncertain parameters*). As already described above, in many cases the uncertain parameter is in fact a (linear) function of the *primitive uncertain parameter* $\boldsymbol{\zeta}$. Note that even though there could be many uncertain parameters in the problem at hand, the number of real or primitive sources of uncertainties is "generally" limited. An important example are the so-called *factor models* in finance, where the uncertain returns of different types of assets are linear functions of a limited

number of economic factors. These economic factors are considered as the primitive uncertain parameters. One of the most famous examples of that is the 3-factor model of Fama and French (1993).

Step 1b (*Determine uncertainty set*). We refer to Section 2.3 for a treatment on natural choices of uncertainty sets.

Step 2 (*Check robustness of nominal solution*). For several applications the nominal optimal solution may already be robust. However, in general using the nominal optimal solution often leads to "severe" infeasibilities. In this step we advocate to do a simulation study to analyze the quality of the nominal solution. If the nominal solution is already robust "enough", then there is of course no need to apply RO. Section 2.10 extensively describes how to do that.

In some applications the constraints are not that strict, and one is more interested in a good "average behavior". Note however that the RO methodology is primarily meant for protecting against the worst case scenario in an uncertainty set. However, often, as a byproduct, the robust solution shows good average behavior, but that is certainly not guaranteed.

If one is interested in a good average behavior, then one may try to use smaller uncertainty sets or use *globalized robust optimization* (GRO); for details on GRO we refer to Ben-Tal et al. (2009, Chapters 3&11).

Step 3a & 3b (*Determine adjustable variables and decision rules*). We discuss several important issues with respect to Step 2, these are listed below.

Reducing extra number of variables. To obtain computationally tractable robust counterpart problems, one often has to use linear decision rules. However, when the number of uncertain parameters is high, the use of linear decision rules may lead to a big increase of the number of variables. Note that these extra variables are added to all constraints that contain adjustable variables. Moreover, when a constraint or the objective in the original problem does not contain uncertain parameters, but does contain adjustable variables, then after substituting the decision rule it will have uncertain parameters, and this will also lead to extra variables in the robust counterpart.

Sometimes, one can choose between a decision rule that is linear in the primitive uncertain parameter $\boldsymbol{\zeta} \in \mathcal{Z}$ or linear in the "general" uncertain parameter $\mathbf{A} \in \mathcal{U}$.

Often the number of primitive uncertain parameters is much smaller, and using them for the decision rule will lead to less variables. In Section 2.4 the advantages and disadvantages of both choices are treated. In many cases we have to restrict the linear decision rule to a subset of the uncertain vector $\boldsymbol{\zeta}$. This is especially the case in multi-period situations. In a production-inventory situation, for example, a linear decision rule in period t can only depend on the known demand of period 1 to $t-1$, since the demand in periods $t, t+1$ and so on is not known yet. This also reduces the number of extra variables.

To further avoid a big increase in the number of variables because of the linear decision rule, one can try to use a subset of the uncertain vector $\boldsymbol{\zeta}$ that is called the "information base". In a production-inventory situation, for example, we may choose a linear decision rule in period t that depends on the known demand of, example given, the last two periods $t-1$ and $t-2$. This reduces the number of variables a lot, and numerical experiments have shown that often the resulting decision rule is almost as good as the full linear one; e.g., see Ben-Tal et al. (2009). By comparing different information bases one could calculate the value of information.

Often an optimization problem contains *analysis* variables. As an example we give the inventory at time t in a production-inventory problem. For such analysis variables we can use a decision rule that depends on all the uncertain parameters, since we do not have to know the value of these analysis variables "here-and-now". The advantage of making analysis variables adjustable is that this may lead to better objective values. The disadvantage of this, however, is the increase of the number of extra variables.

Integer adjustable variables. A parametric decision rule, like the linear one, cannot be used for integer adjustable variables, since we have then to enforce that the decision rule is integer for all $\boldsymbol{\zeta} \in \mathcal{Z}$. In Section 2.5 we propose a new general way for dealing with adjustable integer variables. However, much more research is needed. In Section 2.6 we show that in some cases the integer variables are automatically adjustable.

Quadratic uncertainty. Suppose that we use a quadratic decision rule instead of a linear one. Then, the corresponding robust counterpart is still linear in all the optimization variables, but quadratic in the uncertain parameters. Hence, if the uncertainty set is ellipsoidal, we can use the results from Ben-Tal et al. (2009) to obtain a tractable reformulation. In fact, the final constraint is then a semidefinite programming (SDP) constraint.

Suppose that the situation is not fixed recourse as assumed above, but that $\mathbf{B}$ is also uncertain and linear in ζ. Then using a linear decision rule for y results into quadratic uncertainties. Hence, if the uncertainty set is ellipsoidal, we can use the results from Ben-Tal et al. (2009) to obtain a tractable reformulation. The resulting constraint is again an SDP.

Constraint-wise uncertainty in ARO. We emphasize that if an adjustable variable is used in multiple constraints, those constraints then contain the same set of uncertain parameters, since the adjustable variable is usually a function of all uncertain parameters; see Section 2.2. We have seen that, in RO, without loss of generality we can reformulate the robust problem such that we have constraint-wise uncertainty. However, in ARO, we should first substitute the decision rules for adjustable variables, and then make the uncertainty constraint-wise; but not the other way around, since this may result in incorrect reformulations.

It can be shown that when the uncertainty in the original robust optimization problem is constraint-wise, then the objective values of ARC and RC are the same Ben-Tal et al. (2009). Hence, in such cases using decision rules for adjustable variables does not lead to better objective values. However, there may still be value in using ARC since this may lead to (much) better average behavior; see the numerical example in Section 2.5.

Folding horizon. If one is allowed to reoptimize after each stage in a multi-stage problem, one can of course use adjustable robust optimization in each stage, using that part of the uncertain data that has been revealed. This is called a *folding horizon* (FH) strategy. To compare the ARC FH strategy with the nominal solution, one should also apply a FH strategy to the nominal optimization problem. One could also apply the RC approach in a FH. In many cases this is a good alternative for the ARC approach, e.g., when the ARC approach leads to too large problems. Moreover, RC FH may lead to better solutions than ARC FH; see Section 2.11.

Step 4 (*Formulate robust counterpart*). RO has also to do with modeling. The modeling part is often overlooked in RO applications. An error often made in practice is that the robustness is added to the model after reformulation of the deterministic model. This often leads to solutions that are too conservative. Hence, an important warning is that the robust versions of two equivalent deterministic optimization problems may not be equivalent. We refer to Section 2.7 for a detailed treatment on

these modeling issues.

We also observed that in several applications there are only one or a few uncertain parameters in each constraint, but the uncertainty set is a "joint" region (e.g., ellipsoidal region). Using the constraint-wise interpretation of the RO methodology may be too conservative for such problems, especially in the case where the constraint are not that strict.

It is very important to understand the basic RO concept. What does it mean that RO protects against the worst case scenario? Section 2.9 explains this in more detail.

Step 5 (*Solve RC via tractable reformulation*). If the constraints are linear in the uncertain parameters and in the optimization variables, then there are two ways to derive a tractable reformulation. The first way is the constraint-wise approach by Ben-Tal et al. (2012) that uses Fenchel duality; see Table 2.1 for a summary. The second way is to solve the dual problem of the robust counterpart problem. This approach can handle all compact and convex uncertainty sets; see Gorissen et al. (2012). If the constraints are nonlinear in the uncertain parameter and/or the variables, we refer to Ben-Tal et al. (2012) for deriving tractable robust counterparts. However, we emphasize that for many of such problems it might be not possible to derive tractable robust counterparts.

In Iancu and Trichakis (2013) it is observed that (A)RCs may have multiple optimal solutions. We advice to check whether this is the case, and to use a two-step procedure to find Pareto optimal solutions and to improve on the average behavior; for details see Section 2.5, Iancu and Trichakis (2013), and de Ruiter (2013).

Step 6 (*Solve RC via adversarial approach*). If the robust counterpart cannot be written as or approximated by a tractable reformulation, we advocate to perform the so-called *adversarial approach*. The adversarial approach starts with a finite set of scenarios $S_i \subset \mathcal{Z}_i$ for the uncertain parameter in constraint i. E.g., at the start, S_i only contains the nominal scenario. Then, the robust optimization problem, in which $\mathcal{Z}_i$ is replaced by S_i is solved. If the resulting solution is robust feasible, we have found the robust optimal solution. If that is not the case, we can find a scenario for the uncertain parameter that makes the last found solution infeasible. E.g., we can search for the scenario that maximizes the infeasibility. We add this scenario to S_i, and solve the resulting robust optimization problem, and so on. For a more detailed description, we refer to Bienstock and Özbay (2008). It appeared that this

Table 2.1 – Tractable reformulations for the uncertain constraint $[(\mathbf{a^0} + \mathbf{P\zeta})^\top \mathbf{x} \leq d \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}]$, and h_k^* is the convex conjugate of h_k

Uncertainty	$\mathcal{Z}$	Robust Counterpart	Tractability
Box	$\|\boldsymbol{\zeta}\|_\infty \leq \rho$	$(\mathbf{a^0})^\top \mathbf{x} + \rho \|\mathbf{P}^\top \mathbf{x}\|_1 \leq d$	LP
Ellipsoidal	$\|\boldsymbol{\zeta}\|_2 \leq \rho$	$(\mathbf{a^0})^\top \mathbf{x} + \rho \|\mathbf{P}^\top \mathbf{x}\|_2 \leq d$	CQP
Polyhedral	$\mathbf{D\zeta} + \mathbf{d} \geq 0$	$\begin{cases} (\mathbf{a^0})^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} \leq d \\ \mathbf{D}^\top \mathbf{y} = -\mathbf{P}^\top \mathbf{x} \\ \mathbf{y} \geq \mathbf{0} \end{cases}$	LP
Convex cons.	$h_k(\boldsymbol{\zeta}) \leq 0 \quad \forall k$	$\begin{cases} (\mathbf{a^0})^\top \mathbf{x} + \sum_k u_k h_k^* \left(\frac{\mathbf{w^k}}{u_k}\right) \leq d \\ \sum_k \mathbf{w^k} = \mathbf{P}^\top \mathbf{x} \\ \mathbf{u} \geq \mathbf{0} \end{cases}$	Convex Opt.

simple approach often converges to optimality in a few number of iterations. The advantage of this approach is that solving the robust optimization problem with S_i instead of $\mathcal{Z}_i$ in each iteration, preserves the structure of the original optimization problem. Only constraints of the same type are added, since constraint i should hold for all scenarios in S_i.

We also note that in some cases it may happen that although a tractable reformulation of the robust counterpart can be derived, the size of the resulting problem becomes too big. For such cases the adversarial approach can also be used.

Step 7 (*Check quality of solution*). Since a robustness analysis is extremely important, and in practice one can easily draw wrong conclusions, we extensively describe in Section 2.10 how to perform such an analysis. Frequently stated criticism on RO is that it yields overly pessimistic solutions. Besides performing a wrong robustness analysis, there are several other possible reasons for such criticism. The first is that in the modeling phase one could easily introduce unnecessary pessimism when one does not realize that the robust counterpart of equivalent deterministic problems are not necessarily equivalent. For a detailed explanation on this issue, see Section 2.7. A second reason may be that the constraints that contain uncertain parameters are not that strict as e.g. safety restrictions for the design of a bridge or an airplane. In such cases violating the constraint for some scenarios of the uncertain parameters is not that serious. As it is explained in Step 1b, for those cases one could use the GRO methodology or, alternatively, reduce the size of the uncertainty region. These alternatives can also be used when one is more interested in the average than the

worst case behavior. Finally, there are also cases where indeed the nominal solution is already robust "enough", and where RO does not yield better and more robust solutions. We argue that in practice such a conclusion is already extremely valuable.

2.3 Choosing Uncertainty Set

In this section we describe different possible uncertainty sets and their advantages and disadvantages. Often one wants to make a trade-off between "full" robustness and the size of the uncertainty set: a box uncertainty set that contains the full range of realizations for each component of $\boldsymbol{\zeta}$ is the most robust choice and guarantees that the constraint is never violated, but on the other hand there is only a small chance that all uncertain parameters take their worst case values. This has led to the development of smaller uncertainty sets that still guarantee that the constraint is "almost never" violated. In this thesis, we propose data driven methods, in a similar vein, constructing uncertainty sets using data and probability guarantees is inspired by chance constraints, which are constraints that have to hold with at least a certain probability. Often the underlying probability distribution is not known, and one seeks a distributionally robust solution. One application of RO is to provide a tractable safe approximation of the chance constraint in such cases, i.e. a tractable formulation that guarantees that the chance constraint holds:

$$\text{if } x \text{ satisfies } \mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d \quad \forall \boldsymbol{\zeta} \in \mathcal{U}_\varepsilon, \text{ then } x \text{ also satisfies } \mathbb{P}_{\boldsymbol{\zeta}}(\mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d) \geq 1-\varepsilon.$$

For $\varepsilon = 0$, a chance constraint is a traditional robust constraint. The challenge is to determine the set $\mathcal{Z}_\varepsilon$ for other values of ε. We distinguish between uncertainty sets for uncertain parameters and for uncertain probability vectors.

For uncertain parameters, many results are given in (Ben-Tal et al., 2009, Chapter 2). The simplest case is when the only knowledge about $\boldsymbol{\zeta}$ is that $||\boldsymbol{\zeta}||_\infty \leq 1$. For this case, the box uncertainty set is the only set that can provide a probability guarantee (of $\varepsilon = 0$). When more information becomes available, such as bounds on the mean or variance, or knowledge that the probability distribution is symmetric or unimodal, smaller uncertainty sets become available. Ben-Tal et al. (2009, Table 2.3) list seven of these cases. Probability guarantees are only given when $||\boldsymbol{\zeta}||_\infty \leq 1$, $\mathbb{E}(\boldsymbol{\zeta}) = \mathbf{0}$ and the components of $\boldsymbol{\zeta}$ are independent. We mention the uncertainty sets that are used in practice when box uncertainty is found to be too pessimistic. The first is an ellipsoid (Ben-Tal et al., 2009, Proposition 2.3.1), possibly intersected with a box (Ben-Tal et al., 2009, Proposition 2.3.3):

$$\mathcal{Z}_\varepsilon = \{\boldsymbol{\zeta} : ||\boldsymbol{\zeta}||_2 \leq \Omega \quad ||\boldsymbol{\zeta}||_\infty \leq 1\}, \tag{2.2}$$

where $\varepsilon = \exp(-\Omega^2/2)$. The second is a polyhedral set (Ben-Tal et al., 2009, Proposition 2.3.4), called budgeted uncertainty set or the "Bertsimas and Sim" uncertainty set (Bertsimas and Sim, 2004):

$$\mathcal{Z}_\varepsilon = \{\boldsymbol{\zeta} : ||\boldsymbol{\zeta}||_1 \leq \Gamma \quad ||\boldsymbol{\zeta}||_\infty \leq 1\}, \tag{2.3}$$

where $\varepsilon = \exp(-\Gamma^2/(2L))$. The probability guarantee of the Bertismas and Sim uncertainty set is only valid when the uncertain parameters are independent and symmetrically distributed. A stronger bound is provided in (Bertsimas and Sim, 2004). This set has the interpretation that (integer) Γ controls the number of elements of $\boldsymbol{\zeta}$ that may deviate from their nominal values. (2.2) leads to better objective values for a fixed ε compared to (2.3), but gives rise to a CQP for an uncertain LP while (2.3) results in an LP and is therefore from a computational point of view more tractable.

Bandi and Bertsimas (2012) propose uncertainty sets based on the central limit theorem. When the components of $\boldsymbol{\zeta}$ are independent and identically distributed with mean μ and variance σ^2, the uncertainty set is given by:

$$\mathcal{Z}_\varepsilon = \left\{\boldsymbol{\zeta} : |\sum_{i=1}^{L} \zeta_i - L\mu| \leq \rho\sqrt{n}\sigma\right\},$$

where ρ controls the probability of constraint violation $1 - \varepsilon$. Bandi and Bertsimas also show variations on $\mathcal{Z}_\varepsilon$ that incorporate correlations, heavy tails, or other distributional information. The advantage of this uncertainty set is its tractability, since the robust counterpart of an LP with this uncertainty set is also LP. A disadvantage of this uncertainty set is that it is unbounded for $L > 1$, since one component of $\boldsymbol{\zeta}$ can be increased to an arbitrarily large number (while simultaneously decreasing a different component). This may lead to intractability of the robust counterpart or to trivial solutions. In order to avoid infeasibility, it is necessary to define separate uncertainty sets for each constraint, where the summation runs only over the elements of $\boldsymbol{\zeta}$ that appear in that constraint. Alternatively, it may help to take the intersection of $\mathcal{Z}_\varepsilon$ with a box.

We now focus on uncertain probability vectors. These appear e.g. in a constraint on a risk measure expected value or variance. Ben-Tal et al. (2013) construct uncertainty sets based on ϕ-divergence. The ϕ-divergence between the vectors $\mathbf{p}$ and $\mathbf{q}$ is:

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right),$$

where ϕ is the (convex) ϕ-divergence function; for details on ϕ-divergence, we refer to Pardo (2005). Let $\mathbf{p}$ denote a probability vector and let $\mathbf{q}$ be the vector with observed frequencies when N items are sampled according to $\mathbf{p}$. Under certain regularity conditions,

$$\frac{2N}{\phi''(1)} I_\phi(\mathbf{p}, \mathbf{q}) \xrightarrow{d} \chi^2_{m-1} \text{ as } N \to \infty.$$

This motivates the use of the following uncertainty set:

$$\mathcal{Z}_\varepsilon = \{\mathbf{p} : \mathbf{p} \geq \mathbf{0}, \quad \mathbf{e}^\top \mathbf{p} = 1, \quad \frac{2N}{\phi''(1)} I_\phi(\mathbf{p}, \hat{\mathbf{p}}) \leq \chi^2_{m-1;1-\varepsilon}\},$$

where $\hat{\mathbf{p}}$ is an estimate of $\mathbf{p}$ based on N observations, and $\chi^2_{m-1;1-\varepsilon}$ is the $1-\varepsilon$ percentile of the χ^2 distribution with $m-1$ degrees of freedom. The uncertainty set contains the true $\mathbf{p}$ with (approximate) probability $1-\varepsilon$. Ben-Tal et al. (2013) give many examples of ϕ-divergence functions that lead to tractable robust counterparts.

An alternative to ϕ-divergence is using the Anderson-Darling test to construct the uncertainty set (Ben-Tal et al., 2012, Ex. 15).

We conclude this section by pointing out a mistake that is sometimes made regarding the probability of violation. Sometimes an uncertainty set is constructed such that it contains the true parameter with high probability. Consequently, the constraint holds with the same high probability. However, the probability of constraint satisfaction is much larger than one expects, since the constraint also holds for the "good" realizations of the uncertain parameter outside the uncertainty set. We demonstrate this with a normally distributed $\boldsymbol{\zeta}$ of dimension $L = 10$, where the components are independent, and have mean 0 and variance 1. The singleton $\mathcal{Z}_\varepsilon = \{\mathbf{0}\}$ already guarantees that the uncertain constraint holds with probability 0.5. Let us now construct a set $\mathcal{Z}_\varepsilon$ that contains $\boldsymbol{\zeta}$ with probability 0.5. Since $\boldsymbol{\zeta}^\top \boldsymbol{\zeta} \sim \chi^2_L$, the set $\mathcal{Z}_\varepsilon = \{\boldsymbol{\zeta} : ||\boldsymbol{\zeta}||_2 \leq \sqrt{\chi^2_{L;1-\varepsilon}}\}$ contains $\boldsymbol{\zeta}$ with probability $1-\varepsilon$. For $\varepsilon = 0.5$, $\mathcal{Z}_\varepsilon$ is a ball with radius 9.3 which is indeed much larger than the singleton. Consequently, it provides a much stronger probability guarantee. In order to compute this probability, we first write the explicit chance constraint. Since $(\mathbf{a^0} + \mathbf{P}\boldsymbol{\zeta})^\top \mathbf{x} \leq d$ is equivalent to $(\mathbf{a^0})^\top \mathbf{x} + (\mathbf{P}^\top \mathbf{x})^\top \boldsymbol{\zeta} \leq d$, and since the term $(\mathbf{P}^\top \mathbf{x})^\top \boldsymbol{\zeta}$ follows a normal distribution with mean 0 and standard deviation $\left|\left|\mathbf{P}^\top \mathbf{x}\right|\right|_2$, the chance constraint can explicitly be formulated as $(\mathbf{a^0})^\top \mathbf{x} + z_{1-\varepsilon} \left|\left|\mathbf{P}^\top \mathbf{x}\right|\right|_2 \leq d$, where $z_{1-\varepsilon}$ is the $1-\varepsilon$ percentile of the normal distribution. This is the robust counterpart of the original linear constraint with ellipsoidal uncertainty and a radius of $z_{1-\varepsilon}$. The value $z_{1-\varepsilon} = 9.3$ coincides with $\varepsilon \approx 7.0 \cdot 10^{-21}$. So, while one thinks to construct a set that makes the constraint

hold in 50% of the cases, the set actually makes the constraint hold in almost all cases. To make the chance constraint hold with probability $1-\varepsilon$, the radius of the ellipsoidal uncertainty set is $z_{1-\varepsilon}$ instead of $\sqrt{\chi^2_{L;1-\varepsilon}}$. These only coincide for $L=1$.

2.4 Linearly Adjustable Robust Counterpart: Linear in What?

Tractable examples of decision rules used in ARO are linear (or affine) decision rules (AARC) (Ben-Tal et al., 2009, Chapter 14) or piecewise linear decision rules (Chen et al., 2008); see also Section 2.2. The AARC was introduced by Ben-Tal et al. (2004) as a computationally tractable method to handle adjustable variables. In the following constraint:

$$(\mathbf{a^0} + \mathbf{P}\boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}^\top \mathbf{y} \leq d \qquad \forall \boldsymbol{\zeta} \in \mathcal{Z},$$

$\boldsymbol{y}$ is an adjustable variable whose value may depend on the realization of the uncertain $\boldsymbol{\zeta}$, while $\boldsymbol{b}$ does not depend on $\boldsymbol{\zeta}$ (fixed recourse). There are two different AARCs for this constraint:

AARC 1. $\mathbf{y}$ is linear in $\boldsymbol{\zeta}$ (e.g. see Ben-Tal et al. (2004) and Ben-Tal et al. (2009, Chapter 14)), or

AARC 2. $\mathbf{y}$ is linear in $\mathbf{a^0} + \mathbf{P}\boldsymbol{\zeta}$ (e.g. see Roelofs and Bisschop (2012, Chapter 20.4)).

Note that AARC 2 is as least as conservative as AARC 1, since the linear transformation of $\boldsymbol{\zeta} \mapsto \mathbf{a^0} + \mathbf{P}\boldsymbol{\zeta}$ can only lead to loss of information, and that both methods are equivalent if the linear transformation is injective on $\mathcal{Z}$. The choice for a particular method may be influenced by four factors: (i) the availability of information. An actual decision cannot depend on $\boldsymbol{\zeta}$ if $\boldsymbol{\zeta}$ has not been observed. (ii) The number of variables in the final problem. AARC 1 leads to $|\boldsymbol{\zeta}|$ extra variables compared to the RC, whereas AARC 2 leads to $|\mathbf{a^0}|$ extra variables. (iii) Simplicity for the user. Often the user observes model parameters instead of the primitive uncertainty vector. (iv) For analysis variables one should always use the least conservative method.

The practical issue raised in the first factor (availability of information) has been addressed with a information base matrix $\mathbf{P}$. Instead of being linear in $\boldsymbol{\zeta}$, $\mathbf{y}$ can be made linear in $\mathbf{P}\boldsymbol{\zeta}$. We give one example where uncertain demand is observed. Suppose there are two time periods and three possible scenarios for demand time period one and two, namely $(10,10)^\top$, $(10,11)^\top$ and $(11,11)^\top$. So, the uncertainty set of the demand vector is the convex hull of these scenarios: $\{\boldsymbol{P\zeta} : \boldsymbol{\zeta} \in \mathcal{Z}\}$ where

$\boldsymbol{P}$ is the matrix with the scenarios as columns and $\mathcal{Z} = \Delta^2 = \{\boldsymbol{\zeta} \in \mathbb{R}^3 : \sum_{\ell=1}^{3} \zeta_\ell = 1, \boldsymbol{\zeta} \geq 0\}$ is the standard simplex in $\mathbb{R}^3$. If the observed demand for time period one is 10, it is not possible to distinguish between $\boldsymbol{\zeta} = (1, 0, 0)^\top$ and $\boldsymbol{\zeta} = (0, 1, 0)^\top$. So, a decision for time period two can be modeled either as AARC 1 with $\mathbf{P} = (1, 1, 0)$ or as AARC 2. The latter leads to a decision rule that is easier to interpret, since it directly relates to previously observed demand.

2.5 Adjustable Integer Variables

Ben-Tal et al. (2009, Chapter 14) use parametric decision rules for adjustable continuous variables. However, their novel techniques "generally" cannot be applied for adjustable integer variables. In the literature two alternative approaches have been proposed. Bertsimas and Georghiou (2013) introduced an iterative method to treat adjustable *binary* variables as piecewise constant functions. The approach by Bertsimas and Caramanis (2010) is different and is based on splitting the uncertainty region into smaller subsets, where each subset has its own binary decision variable (see also Vayanos et al. (2011)). In this section, we briefly show this last method to treat adjustable integer variables, and show how the average behavior can be improved. We use the following notation for the general RC problem:

$$\begin{aligned} \text{(RC1)} \quad & \max_{\mathbf{x},\mathbf{y},\mathbf{z}} \; c(\mathbf{x}, \mathbf{y}, \mathbf{z}) \\ & \text{s.t.} \; \mathbf{A}(\boldsymbol{\zeta})\, \mathbf{x} + \mathbf{B}(\boldsymbol{\zeta})\, \mathbf{y} + \mathbf{C}(\boldsymbol{\zeta})\, \mathbf{z} \leq \mathbf{d}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^{n_1}$ and $\mathbf{y} \in \mathbb{Z}^{n_2}$ are "here and now" variables, i.e., decisions on them are made before the uncertain parameter $\boldsymbol{\zeta}$, contained in the uncertainty set $\mathcal{Z} \subseteq \mathbb{R}^L$, is revealed; $\mathbf{z} \in \mathbb{Z}^{n_3}$ is a "wait and see" variable, i.e., the decision on $\mathbf{z}$ is made after observing (part of) the value of the uncertain parameter. $\mathbf{A}(\boldsymbol{\zeta}) \in \mathbb{R}^{m_1 \times n_1}$ and $\mathbf{B}(\boldsymbol{\zeta}) \in \mathbb{R}^{m_2 \times n_2}$ are the uncertain coefficient matrices of the "here and now" variables. Notice that the integer "wait and see" variable $\mathbf{z}$ has an uncertain coefficient matrix $\mathbf{C}(\boldsymbol{\zeta}) \in \mathbb{R}^{m_3 \times n_3}$. So, unlike the "classic" parametric method, this approach can handle uncertainties in the coefficients of the integer "wait and see" variables. For the sake of simplicity, we assume that the uncertain coefficient matrices to be linear in $\boldsymbol{\zeta}$ and, without loss of generality, $c(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is the certain linear objective function.

To model the *adjustable* RC (ARC) with integer variables, we first divide the given uncertainty set $\mathcal{Z}$ into m disjoint, excluding the boundaries, subsets ($\mathcal{Z}_i$, $i = 1, \ldots, m$):

$$\mathcal{Z} = \bigcup_{i \in \{1,\ldots,m\}} \mathcal{Z}_i,$$

and we introduce additional integer variables $\mathbf{z_i} \in \mathbb{Z}^{n_3}$ $(i = 1, \ldots, m)$ that model the decision in $\mathcal{Z}_i$. Then, we replicate the uncertain constraint and the objective function in (RC1) for each $\mathbf{z_i}$ and the uncertainty set $\mathcal{Z}_i$ as follows:

$$\begin{aligned}
\text{(ARC1)} \quad \max_{\mathbf{x},\mathbf{y},\mathbf{Z},t} \quad & t \\
\text{s.t.} \quad & c(\mathbf{x}, \mathbf{y}, \mathbf{z_i}) \geq t && \forall i \in \{1, \ldots, m\} \qquad (2.4)\\
& \mathbf{A}(\boldsymbol{\zeta})\, \mathbf{x} + \mathbf{B}(\boldsymbol{\zeta})\, \mathbf{y} + \mathbf{C}(\boldsymbol{\zeta})\, \mathbf{z_i} \leq \mathbf{d} && \forall \boldsymbol{\zeta} \in \mathcal{Z}_i, \forall i \in \{1, \ldots, m\}.
\end{aligned}$$

Note that (ARC1) is more flexible than the non-adjustable RC (RC1) in selecting the values of integer variables, since it has a specific decision $\mathbf{z_i}$ for each subset $\mathcal{Z}_i$. Therefore, (ARC1) yields a robust optimal objective that is at least as good as (RC1).

Pareto efficiency. Iancu and Trichakis (2013) discovered that "the inherent focus of RO on optimizing performance only under worst case outcomes might leave decisions un-optimized in case a non worst case scenario materialized". Therefore, the "classical" RO framework might lead to Pareto inefficiencies; i.e., an alternative robust optimal solution may guarantee an improvement in the objective for (at least) a scenario without deteriorating it in other scenarios.

Pareto efficiency is also an issue in (ARC1) that coincides with the *worst case* objective value among m objective functions associated with the subsets. Henceforth, we must take into account the individual performance of the m subsets to have a better understanding of the general performance of (ARC1). To find Pareto efficient robust solutions, Iancu and Trichakis propose reoptimizing the slacks of "important" constraints, i.e., defined by a value vector, by fixing the robust optimal objective value of the classical RO problem that is initially optimized; for details on pareto efficiency in robust linear optimization we refer to Iancu and Trichakis (2013). Following a similar approach, we apply a reoptimization procedure to improve the average performance of (ARC1). More precisely, we first solve (ARC1) and find the optimal objective t^*. Then, we solve the following problem:

$$\begin{aligned}
\text{(re-opt)} \quad \max_{\mathbf{x},\mathbf{y},\mathbf{Z},\mathbf{t}} \quad & \sum_{i \in \{1,\ldots,m\}} t_i \\
\text{s.t.} \quad & t_i \geq t^* && \forall i \in \{1, \ldots, m\} \\
& c(\mathbf{x}, \mathbf{y}, \mathbf{z_i}) \geq t_i && \forall i \in \{1, \ldots, m\} \\
& \mathbf{A}(\boldsymbol{\zeta})\, \mathbf{x} + \mathbf{B}(\boldsymbol{\zeta})\, \mathbf{y} + \mathbf{C}(\boldsymbol{\zeta})\, \mathbf{z_i} \leq \mathbf{d} && \forall \boldsymbol{\zeta} \in \mathcal{Z}_i, \forall i \in \{1, \ldots, m\},
\end{aligned}$$

that optimizes (i.e., maximizing) the slacks in (2.4), while the worst case objective value t^* remains the same. Note that t_i's are the additional variables associated with

the objectives values of the subsets; (re-opt) mimics a multi-objective optimization problem that assigns equal weights to each objective, and finds Pareto efficient robust solutions.

Example

Here we compare the optimal objective values of (RC1), (ARC1), and (ARC1) with (re-opt) via a toy example. For the sake of exposition, we exclude continuous variables in this example. The non-adjustable RC is given as follows:

$$\begin{aligned} \max_{(w,\mathbf{z})\in\mathbb{Z}^3_+} \quad & 5w + 3z_1 + 4z_2 \\ \text{s.t.} \quad & (1+\zeta_1+2\zeta_2)w + (1-2\zeta_1+\zeta_2)z_1 + (2+2\zeta_1)z_2 \le 18 \quad \forall \boldsymbol{\zeta} \in \text{Box} \\ & (\zeta_1+\zeta_2)w + (1-2\zeta_1)z_1 + (1-2\zeta_1-\zeta_1)z_2 \le 16 \qquad \forall \boldsymbol{\zeta} \in \text{Box}, \end{aligned} \tag{2.5}$$

where $\text{Box} = \{\zeta : -1 \le \zeta_1 \le 1, -1 \le \zeta_2 \le 1\}$ is the given uncertainty set, and w, z_1, and z_2 are nonnegative integer variables. In addition, we assume that z_1 and z_2 are adjustable on ζ_1; i.e., the decision on these variables is made after ζ_1 is being observed. Next, we divide the uncertainty set into two subsets:

$$\begin{aligned} \mathcal{Z}_1 &= \{(\zeta_1,\zeta_2) : -1 \le \zeta_1 \le 0, -1 \le \zeta_2 \le 1\} \\ \mathcal{Z}_2 &= \{(\zeta_1,\zeta_2) : 0 \le \zeta_1 \le 1, -1 \le \zeta_2 \le 1\}. \end{aligned}$$

Then ARC of (2.5) is:

$$\begin{aligned} \text{(Ex:ARC)} \quad \max_{t,w,\mathbf{Z}} \quad & t \\ \text{s.t.} \quad & 5w + 3z_1^i + 4z_2^i \ge t \qquad \forall i \\ & (1+\zeta_1+2\zeta_2)w + (1-2\zeta_1+\zeta_2)z_1^i + (2+2\zeta_1)z_2^i \le 18 \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_i, \forall i \\ & (\zeta_1+\zeta_2)w + (1-2\zeta_1)z_1^i + (1-2\zeta_1-\zeta_1)z_2^i \le 16 \qquad \forall \boldsymbol{\zeta} \in \mathcal{Z}_i, \forall i, \end{aligned}$$

where $t \in \mathbb{R}$, $w \in \mathbb{Z}_+$, $\mathbf{Z} \in \mathbb{Z}_+^{2\times m}$, and $m = 2$ since we have two subsets. Table 2.2 presents the optimal solutions of RC and ARC problems.

Table 2.2 – RC vs ARC

Method	Obj.	w	$\mathbf{z}$
RC	29	1	$(z_1, z_2) = (4, 3)$
ARC	31	0	$(z_1^1, z_2^1, z_1^2, z_2^2) = (0, 8, 5, 4)$

The numerical results show that using the adjustable reformulation we improve the objective value of the non-adjustable problem by 7%. On the other hand, if we assume that z_1 and z_2 are adjustable on ζ_2 (but not on ζ_1), and we modify the uncertainty subsets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ accordingly, then RC and ARC yield the same objective 29. This shows that the value of information of ζ_1 is higher than that of ζ_2.

Next we compare the average performance of ARC and the second stage optimization problem (re-opt) that is given by:

$$\begin{aligned}
&\max_{\mathbf{t},w,\mathbf{Z}} \sum_{i\in\{1,\dots,m\}} t_i \\
&\text{s.t. } 5w + 3z_1^i + 4z_2^i \geq t_i, \quad t_i \geq t^* && \forall i \\
&\qquad (1+\zeta_1+2\zeta_2)w + (1-2\zeta_1+\zeta_2)z_1^i + (2+2\zeta_1)z_2^i \leq 18 && \forall \boldsymbol{\zeta} \in \mathcal{Z}_i, \forall i \\
&\qquad (\zeta_1+\zeta_2)w + (1-2\zeta_1)z_1^i + (1-2\zeta_1-\zeta_2)z_2^i \leq 16 && \forall \boldsymbol{\zeta} \in \mathcal{Z}_i, \forall i,
\end{aligned}$$

where $\mathbf{t} \in \mathbb{R}^m$. For changing the number of subsets, we again split the uncertainty sets $(\mathcal{Z}_i, i = 1, \dots, m)$ on ζ_1 but not on ζ_2. The numerical results are presented in Table 2.3.

Table 2.3 – ARC vs re-opt for varying number of subsets

	Worst Case Obj. Values per Subset		W.-C. Average	
# Subsets	ARC	re-opt	ARC	re-opt
1	29	29	29	29.0
2	(32, 31*)	(34, 31*)	31.5	32.5
3	(33, 30*, 32)	(49, 30*, 35)	31.6	38.0
4	(33, 31*, 32, 32)	(64, 34, 31*, 54)	32	45.7
5	(33, 30*, 30*, 32, 32)	(80, 40, 30*, 33, 66)	31.4	49.8
8	(32, 32, 32, 34, 31*, 33, 33, 33)	(128, 64, 40, 34, 31* 36, 54, 108)	32.5	61.8
10	(32, 32, 32, 32, 34, 31*, 33, 33, 33, 33)	(160, 80, 52, 40, 34, 31*, 33, 45, 66, 135)	32.5	64.3

$(*)$ denotes the worst case (w.-c.) objective value over all subsets

The first column of the table presents the number of subsets used in ARC, and we assume that the domain of ζ_1 is divided into equally sized intervals (e.g., if the number of subsets is three, then the intervals are $-1 \leq \zeta_1 \leq -0.33, -0.33 \leq \zeta_1 \leq 0.33$, and $0.33 \leq \zeta_1 \leq 1$). The second column reports objective values of the subproblems associated with the subsets in ARC. The third column presents the objective values

of the subproblems when we apply (re-opt). The fourth and fifth columns show the averages of the results in columns two and three. As anticipated, intra row comparisons show that ARC and (re-opt) yield the same worst case performance for a fixed number of subsets, and (re-opt) yields significantly better average performances than ARC. Moreover, the average performance improves with the number of subsets. Notice that the average performance of the RC solution is not reported in Table 2.3 because it has the same average performance, that is 29, for any given number of subsets. Nevertheless, it is important to see the significant average objective value improvement made by ARC with (re-opt) for the "fixed" performance of the RC. Last but not least, the optimal objective value 31, which is obtained when the number of subsets is two, four, eight and ten in Table 2.3, is the global optimal of the ARC; for details on optimality see the following section where this example will be revisited.

Optimality

To quantify how far is the optimal objective value (t^*) of (ARC1) from that of the best possible solution, we need to define an efficient lower bound (or an upper bound for a maximization problem) for the best objective. One way of finding such a bound is solving (ARC1) by defining an adjustable variable for each scenario, and scenarios are associated with a finite subset (denoted by $\hat{\mathcal{Z}}$) of the uncertainty set $\mathcal{Z}$ (Hadjiyiannis et al., 2011; Bertsimas and Georghiou, 2013; Postek, 2013). The optimal objective value of such a formulation is always a lower bound for the best possible objective value, since we optimize adjustable variables for each unique scenario separately and use a finite subset $\hat{\mathcal{Z}}$ that is less conservative (or performs the same in the worst case) than the original uncertainty set $\mathcal{Z}$. More precisely, the lower bound problem is given as follows:

$$
\begin{aligned}
\text{(BRC)} \quad \min_{\mathbf{x},\mathbf{y},\mathbf{z}^{(\zeta)},t^{\text{lb}}} \quad & t^{\text{lb}} \\
\text{s.t.} \quad & c(\mathbf{x},\mathbf{y},\mathbf{z}^{(\zeta)}) \leq t^{\text{lb}} && \forall \zeta \in \hat{\mathcal{Z}} \\
& \mathbf{A}(\boldsymbol{\zeta})\,\mathbf{x} + \mathbf{B}(\boldsymbol{\zeta})\,\mathbf{y} + \mathbf{C}(\boldsymbol{\zeta})\,\mathbf{z}^{(\zeta)} \leq \mathbf{d} && \forall \zeta \in \hat{\mathcal{Z}}
\end{aligned}
$$

where $\hat{\mathcal{Z}}$ is a finite subset of $\mathcal{Z}$, as explained above. Now the question that has to be answered is: how to construct $\hat{\mathcal{Z}}$ efficiently? Postek (2013) proposes to first find the optimal solution of (ARC1) for a given number of subsets, and then formulating the set of worst case uncertain parameters for the left-hand sides in active constraints to construct $\hat{\mathcal{Z}}$. For additional details on improving the lower bound we refer to Postek (2013, Chapter 4.2).

Example (Ex:ARC) revisited. The solution of (Ex:ARC) for two subsets (i.e., $m = 2$) is given in the second row of Table 2.2. The associated finite "worst case" subset for this solution is $\hat{\mathcal{Z}} = \{(0, 1), (0, -1)\}$, and the upper bound for the best possible worst case objective is $t^{\text{ub}} = 31$ (this is obtained by solving the upper bound reformulation of (BRC) for $\hat{\mathcal{Z}}$). Therefore, the optimal objective value of (Ex:ARC) is bounded above by 31 for any given number of subsets; but since we find the same objective value for $m = 2$ we can conclude that 31 is the global optimal value.

Tractability

It is important to point out that our adjustable reformulation and the "non-adjustable" RC have the same "general" mathematical complexity, but the adjustable reformulation increases the number of variables and constraints by a factor m (the number of subsets), so that if the number of integer variables is high (say a few hundreds) then the resulting adjustable RC may be intractable. Dividing the main uncertainty set $\mathcal{Z}$ into more subsets $\mathcal{Z}_i$ may improve the objective value by giving more freedom in making adjustable decisions, but the decision maker should make the tradeoff between optimality and computational complexity.

2.6 Binary Variables in Big-M-Constraints are Automatically Adjustable

Often integer variables correspond to strategic here-and-now decisions, and then there is no need to make them adjustable. In this section we show that for an important class of $0, 1$ variables that are wait-and-see there is also no need to make them adjustable.

Suppose y is a $0, 1$ variable that is associated to a continuous variable x in such a way that:

$$y = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Such $0, 1$ variables are often used, e.g., in supply chain models to model whether a facility is opened or not. Now suppose that both y and the continuous variable x are adjustable with respect to the uncertain parameter ζ. Let us use a linear decision rule for x:

$$x = u + \mathbf{v}^\top \boldsymbol{\zeta},$$

where $u \in \mathbb{R}$, and $\mathbf{v} \in \mathbb{R}^L$ are the coefficients of the linear decision rule, and we do not use a linear decision rule for y, although y is adjustable. For the optimal solution of the corresponding robust optimization problem, we either have (i) $u = 0$ and $\mathbf{v} = \mathbf{0}$ (i.e. $x = 0$), or (ii) $u \neq 0$ or $\mathbf{v} \neq \mathbf{0}$. In case (i) we get $y = 0$, and in case (ii) we get $y = 1$. Hence, the only problematic situation is when $u + \mathbf{v}^\top \boldsymbol{\zeta} = 0$ in case (ii), since then the optimal y should be 0 and not 1. Note however that the probability that $\zeta \in \mathcal{U}$ is such that $u + \mathbf{v}^\top \boldsymbol{\zeta} = 0$ is zero, unless a follows a discrete distribution or $u^* + (\mathbf{v}^*)^\top \boldsymbol{\zeta} = 0$, with u^* and $\mathbf{v}^*$ the optimal values for u and $\mathbf{v}$, is part of the description of $\mathcal{Z}$. Also observe that when $u = 0$ and $\mathbf{v} = \mathbf{0}$, we will automatically get $y = 0$, since the constraint $x \leq My$, with M a big number, is one of the constraints of the original problem and the objective is to minimize costs. A more efficient formulation would be:

$$u \leq My, \quad u \geq -My, \quad \mathbf{v} \leq My\mathbf{1}, \quad \mathbf{v} \geq -My\mathbf{1},$$

where $\mathbf{1}$ is the all one vector. We conclude that there is no need to make y adjustable, i.e. the final optimal linear decision rule,

$$y = \begin{cases} 1 & \text{if} \quad u + \mathbf{v}^\top \boldsymbol{\zeta} > 0 \\ 0 & \text{if} \quad u + \mathbf{v}^\top \boldsymbol{\zeta} = 0, \end{cases}$$

can be obtained by only using linear decision rules for x (and not for y). Note that this conclusion depends on the chosen class of decision rules. Suppose that we would have used piecewise linear decision rules, then we should also make y adjustable. One way to do that is to define a different y value for each interval of the piecewise linear decision rule.

Example. Let us consider the following problem. There are two possible production centers and together they have to produce at least ζ. Production costs per unit are 1 and 3, respectively for production center 1 and 2. Fixed costs for opening the centers are 40 and 10, respectively for center 1 and 2. The mathematical formulation is:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & 40y_1 + 10y_2 + x_1 + 3x_2 \\ \text{s.t.} \quad & x_1 + x_2 \geq \zeta \\ & x_1 \leq My_1 \\ & x_2 \leq My_2 \\ & x_1, x_2 \geq 0 \\ & y_1, y_2 \in \{0, 1\}, \end{aligned}$$

in which $M > 0$ is a big number. Now suppose that ζ is uncertain, with uncertainty interval $[10, 30]$, and both x and y are wait-and-see variables. Although y is adjustable

we only use linear decision rules for x, and solve the following adjustable robust optimization problem:

$$\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \max_{\zeta\in[10,30]} \quad & 40y_1 + 10y_2 + u_1 + v_1\zeta + 3(u_2 + v_2\zeta) & \\
\text{s.t.} \quad & u_1 + v_1\zeta + u_2 + v_2\zeta \geq \zeta & \forall\zeta \in [10,30] \\
& u_1 + v_1\zeta \leq My_1 & \forall\zeta \in [10,30] \\
& u_2 + v_2\zeta \leq My_2 & \forall\zeta \in [10,30] \\
& u_1 + v_1\zeta \geq 0 & \forall\zeta \in [10,30] \\
& u_2 + v_2\zeta \geq 0 & \forall\zeta \in [10,30] \\
& y_1, y_2 \in \{0,1\}. &
\end{aligned}$$

The optimal solution of this problem is:

$$\begin{aligned}
& x_1 = \zeta \quad (u_1 = 0, v_1 = 1) \\
& x_2 = 0 \quad (u_2 = 0, v_2 = 0) \\
& y_1 = 1 \\
& y_2 = 0,
\end{aligned}$$

which indeed can not be improved by using decision rules for y. Hence, indeed there is no need to make y adjustable. Now suppose the uncertainty interval is $[0, 30]$, then the optimal linear decision rule is as above. However, now it can happen that $\zeta = 0$, in which case y_1 should be 0 instead of 1. Hence the final optimal decision rule is:

$$\begin{aligned}
& x_1 = \zeta \\
& x_2 = 0 \\
& y_1 = \begin{cases} 1 & \text{if} \quad \zeta > 0 \\ 0 & \text{if} \quad \zeta = 0 \end{cases} \\
& y_2 = 0.
\end{aligned}$$

2.7 Robust Counterparts of Equivalent Deterministic Problems are not Necessarily Equivalent

In this section we show that the robust counterparts of equivalent deterministic problems are not always equivalent. The message in this section is thus that one has to be careful with reformulating optimization problems, since the corresponding robust counterparts may not be the same.

Let us start with a few simple examples. The first one is similar to the example in Ben-Tal et al. (2009, p. 13). Consider the following constraint:

$$(2 + \zeta)x_1 \leq 1,$$

where ζ is an (uncertain) parameter. This constraint is equivalent to:

$$\begin{cases} (2+\zeta)x_1 + s = 1 \\ s \geq 0. \end{cases}$$

However, the robust counterparts of these two constraint formulations, i.e.

$$(2+\zeta)x_1 \leq 1 \quad \forall \zeta \ : \ |\zeta| \leq 1, \tag{2.8}$$

and

$$\begin{cases} (2+\zeta)x_1 + s = 1 \quad \forall \zeta \ : \ |\zeta| \leq 1 \\ s \geq 0, \end{cases} \tag{2.9}$$

in which the uncertainty set for ζ is the set $\{\zeta : |\zeta| \leq 1\}$, are not equivalent. It can easily be verified that the feasible set for robust constraint (2.8) is: $x_1 \leq 1/3$, while for the robust constraint (4.22) this is $x_1 = 0$. The reason why (2.8) and (4.22) are not equivalent is that by adding the slack variable, the inequality becomes an equality that has to be satisfied for all values of the uncertain parameter, which is very restrictive. The general message is therefore: *do not introduce slack variables in uncertain constraints, unless they are adjustable like in Kuhn et al. (2011), and avoid uncertain equalities.*

Another example is the following constraint:

$$|x_1 - \zeta| + |x_2 - \zeta| \leq 2,$$

which is equivalent to:

$$\begin{cases} y_1 + y_2 \leq 2 \\ y_1 \geq x_1 - \zeta \\ y_1 \geq \zeta - x_1 \\ y_2 \geq x_2 - \zeta \\ y_2 \geq \zeta - x_2. \end{cases}$$

However, the robust versions of these two formulations, namely:

$$|x_1 - \zeta| + |x_2 - \zeta| \leq 2 \quad \forall \ \zeta \ : \ |\zeta| \leq 1, \tag{2.10}$$

and:

$$\begin{cases} y_1 + y_2 \leq 2 \\ y_1 \geq x_1 - \zeta \quad \forall \zeta \ : \ |\zeta| \leq 1 \\ y_1 \geq \zeta - x_1 \quad \forall \zeta \ : \ |\zeta| \leq 1 \\ y_2 \geq x_2 - \zeta \quad \forall \zeta \ : \ |\zeta| \leq 1 \\ y_2 \geq \zeta - x_2 \quad \forall \zeta \ : \ |\zeta| \leq 1, \end{cases} \tag{2.11}$$

are not equivalent. Indeed, it can easily be checked that the set of feasible solutions for (2.10) is $(\theta, -\theta)$, $-1 \leq \theta \leq 1$, but the only feasible solution for (2.11) is $\mathbf{x} = (0, 0)$. The reason for this is that in (2.11) the uncertainty is split over several constraints, and since the concept of RO is constraint-wise, this leads to different problems, and thus different solutions. The following linear optimization reformulation, however, is equivalent to (2.10):

$$\begin{cases} x_1 - \zeta + x_2 - \zeta \leq 2 & \forall\, \zeta \;:\; |\zeta| \leq 1 \\ x_1 - \zeta + \zeta - x_2 \leq 2 & \forall\, \zeta \;:\; |\zeta| \leq 1 \\ \zeta - x_1 + x_2 - \zeta \leq 2 & \forall\, \zeta \;:\; |\zeta| \leq 1 \\ \zeta - x_1 + \zeta - x_2 \leq 2 & \forall\, \zeta \;:\; |\zeta| \leq 1. \end{cases} \tag{2.12}$$

The general rule therefore is: *do not split the uncertainty in one constraint over more constraints, unless the uncertainty is disjoint.* In particular do not use "definition variables" if this leads to such a splitting of the uncertainty.

In the remainder we give a general treatment of some often used reformulation tricks to reformulate nonlinear problems into linear ones, and discuss whether the robust counterparts are equivalent or not.

- **Maximum function.** Consider the following constraint:

  $$\mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + \max_k \mathbf{b_k}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \mathcal{Z},$$

  where $\boldsymbol{\zeta} \in \mathcal{Z}$ is the uncertain parameter, and $\mathbf{a}(\boldsymbol{\zeta})$, $\mathbf{b_k}(\boldsymbol{\zeta})$, and $d(\boldsymbol{\zeta})$ are parameters that depend linearly on $\boldsymbol{\zeta}$. The incorrect reformulation for this constraint is:

  $$\begin{cases} \mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + z \leq d(\boldsymbol{\zeta}) & \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ z \geq \mathbf{b_k}(\zeta)^\top \mathbf{x} & \forall k, \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{cases}$$

  since the uncertainty is split over more constraints. The correct reformulation is:

  $$\mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b_k}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d(\boldsymbol{\zeta}) \quad \forall k, \forall \boldsymbol{\zeta} \in \mathcal{Z}.$$

  Note that in many cases we have "a sum of max":

  $$\mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + \sum_i \max_k \mathbf{b_{ik}}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}.$$

Important examples that contain such constraints are production-inventory problems. We refer to Gorissen and den Hertog (2012) for an elaborate treatment on exact and approximate reformulations of such constraints.

- **Absolute value function.** Note that $|x| = \max\{x, -x\}$, and hence this is a special case of the max function, treated above.

- **Linear fractional program.** Consider the following robust linear fractional problem:

$$\begin{cases} \min\limits_{\mathbf{x}} & \max_{\boldsymbol{\zeta}\in\mathcal{Z}} \frac{\alpha(\boldsymbol{\zeta})+\mathbf{c}(\boldsymbol{\zeta})^\top \mathbf{x}}{\beta(\boldsymbol{\zeta})+\mathbf{d}(\boldsymbol{\zeta})^\top \mathbf{x}} \\ \text{s.t.} & \sum\limits_j a_{ij}x_j \geq b_i \quad \forall i \\ & \mathbf{x} \geq \mathbf{0}, \end{cases} \tag{2.13}$$

where $\boldsymbol{\alpha}(\boldsymbol{\zeta})$, $\mathbf{c}(\boldsymbol{\zeta})$, $\boldsymbol{\beta}(\boldsymbol{\zeta})$, and $\mathbf{d}(\boldsymbol{\zeta})$ are parameters that depend linearly on $\boldsymbol{\zeta}$. Moreover, we assume that $\beta(\boldsymbol{\zeta})+\mathbf{d}(\boldsymbol{\zeta})^\top \mathbf{x} > 0$, for all feasible $\mathbf{x}$ and for all $\boldsymbol{\zeta} \in \mathcal{Z}$. For the non-robust version one can use the Charnes-Cooper transformation that is proposed by Charnes and Cooper (1962) to obtain an equivalent linear optimization problem. However, if we apply this transformation to the robust version, we obtain:

$$\begin{cases} \min\limits_{\mathbf{y},t} & \max\limits_{\boldsymbol{\zeta}\in\mathcal{Z}} \alpha(\boldsymbol{\zeta})t + \mathbf{c}(\boldsymbol{\zeta})^\top \mathbf{y} & \\ \text{s.t.} & \beta(\boldsymbol{\zeta})t + \mathbf{d}(\boldsymbol{\zeta})^\top \mathbf{y} = 1 & \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ & \sum\limits_j a_{ij}y_j \geq b_i t & \forall i \\ & \mathbf{y} \geq \mathbf{0},\ t \geq 0, & \end{cases}$$

which is not equivalent to (2.13) since the uncertainty in the original objective is now split over the objective and a constraint. A better way to deal with such problems is to solve the robust linear problem

$$\begin{cases} \min\limits_{\mathbf{x}} & \max\limits_{\boldsymbol{\zeta}\in\mathcal{Z}} \left[\alpha(\boldsymbol{\zeta}) + \mathbf{c}(\boldsymbol{\zeta})^\top \mathbf{x} - \lambda \left(\beta(\boldsymbol{\zeta}) + \mathbf{d}(\boldsymbol{\zeta})^\top \mathbf{x}\right)\right] \\ \text{s.t.} & \sum\limits_j a_{ij}x_j \geq b_i \\ & \mathbf{x} \geq \mathbf{0}, \end{cases}$$

for a fixed value of λ, and then find the minimal value of λ for which this optimization problem still has a non positive optimal value. One can use for example binary search on λ to do this. For a more detailed treatment of robust fractional problems we refer to Gorissen (2013).

- **Product of binary variables.** Suppose that a robust constraint contains a product of binary variables, say xy, with $x, y \in \{0, 1\}$. Then one can use the standard way to linearize this:

$$\begin{cases} z \leq x \\ z \leq y \\ z \geq x + y - 1 \\ z \geq 0, \end{cases}$$

and replace xy with z. One can use this reformulation since the added constraints do not contain uncertain parameters.

- **Product of binary and continuous variable.** A product of a binary and a continuous variable that occurs in a robust constraint can also be reformulated in linear constraints, in a similar way as above. However, note that in the following robust constraint:

$$\mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + z\mathbf{b}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \mathcal{Z},$$

where $z \in \{0, 1\}$, one cannot use the standard trick:

$$\begin{cases} \mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + zy \leq d(\boldsymbol{\zeta}) & \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ y \geq \mathbf{b}(\boldsymbol{\zeta})^\top \mathbf{x} & \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{cases} \tag{2.14}$$

and then linearize zy. This is not possible since in (2.14) the uncertainty is split over different constraints. A correct reformulation is:

$$\begin{cases} \mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d(\boldsymbol{\zeta}) + M(1 - z) & \boldsymbol{\zeta} \in \mathcal{Z} \\ \mathbf{a}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d(\boldsymbol{\zeta}) + Mz & \boldsymbol{\zeta} \in \mathcal{Z}. \end{cases} \tag{2.15}$$

- K **out of** N **constraints should be satisfied.** Suppose the restriction is that at least K out of the N robust constraints

$$\mathbf{a_i}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d_i(\boldsymbol{\zeta}) \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \tag{2.16}$$

should be satisfied, where $i \in \{1, \ldots, N\}$. Then one can use the standard way

$$\begin{cases} \mathbf{a_i}(\boldsymbol{\zeta})^\top \mathbf{x} \leq d_i(\boldsymbol{\zeta}) + M(1 - z_i) & \forall \boldsymbol{\zeta} \in \mathcal{Z}, \forall i \\ \sum_i z_i \geq K & \\ z_i \in \{0, 1\} & \forall i, \end{cases}$$

where M is a sufficiently big number. However, if the restriction is that $\forall \boldsymbol{\zeta} \in \mathcal{Z}$ at least K out of the N constraints should be satisfied (notice the difference with (2.16)), then the above constraint-wise formulation is not equivalent and is overly conservative. We do not see how to model such a constraint correctly. Maybe an adversarial approach could be used for such constraints.

- **If-then constraint.** Since an "if-then constraint" can be modeled as an at least 1 out of 2 constraints, the above remarks hold.

Up to now we only described linear optimization examples. Similar examples can be given for conic and nonlinear optimization. In Lobo et al. (1998) for example, many optimization problems are given that can be modeled as conic quadratic programming problems. However, for many of them it holds that the corresponding robust counterparts are not the same. This means that if an optimization problem is conic quadratic representable, then the robust counterparts are not automatically the same, and hence in such cases the robust optimization techniques for CQP cannot be used.

2.8 How to Deal with Equality Constraints?

Equality constraints containing uncertain parameters should be avoided as much as possible, since often such constraints restrict the feasibility region drastically or even lead to infeasibility. Therefore, the advice is: *do not use slack variables unless they are adjustable, since using slack variables leads to equality constraints*; see Ben-Tal et al. (2009, Chapter 2). However, equality constraints containing uncertain parameters cannot always be avoided. There are several ways to deal with such uncertain equality constraints:

- In some cases it might be possible to convert the equality constraints into inequality constraints. An illustrating example is the transportation problem: the demand constraints can either be formulated as equality constraints or as inequality constraints. The structure of the problem is such that at optimality these inequalities are tight.

- The equality constraints can be used to eliminate variables. This idea is mentioned in Ben-Tal et al. (2009). However, several questions arise. First of all, after elimination of variables and after the resulting problem has been solved, it is unclear which values to take for the eliminated variables, since they also depend on the uncertain parameters. This is no problem if the eliminated

variables are adjustable variables or analysis variables, since there is no need to know their optimal values. A good example is the production-inventory problem for which one can easily eliminate the analysis variables indicating the inventory in different time periods. See e.g. Ben-Tal et al. (2009). Secondly, suppose the coefficients with respect to the variables that will be eliminated contain uncertain parameters. Eliminating such variables leads to problems that contain non-linear uncertainty, which are much more difficult to solve. To illustrate this, let us consider the following two constraints of an optimization problem:

$$\zeta_1 x_1 + x_2 + x_3 = 1, \quad x_1 + x_2 + \zeta_2 x_3 \leq 5,$$

in which ζ_1 and ζ_2 are uncertain parameters. Suppose that x_1, x_2 and x_3 are all adjustable in ζ_1. Then there are three options for elimination:

1. *Elimination of* x_1. Let us assume that $\zeta_1 = 0$ is not in the uncertainty set. By substituting $x_1 = (1 - x_2 - x_3)/\zeta_1$ the inequality becomes:

$$\left(1 - \frac{1}{\zeta_1}\right) x_2 + \left(\zeta_2 - \frac{1}{\zeta_1}\right) x_3 \leq 5 - \frac{1}{\zeta_1}.$$

   The disadvantage of eliminating x_1 is thus that the uncertainty in the inequality becomes nonlinear.

2. *Elimination of* x_2. By substituting $x_2 = 1 - \zeta_1 x_1 - x_3$ the inequality becomes:

$$(1 - \zeta_1)x_1 + (\zeta_2 - 1)x_3 \leq 4,$$

   which is linear in the uncertain parameters.

3. *Elimination of* x_3. By substituting $x_3 = 1 - \zeta_1 x_1 - x_2$ the inequality becomes:

$$(1 - \zeta_1\zeta_2)x_1 + (1 - \zeta_2)x_2 \leq 5 - \zeta_2,$$

   which is nonlinear in the uncertain parameters. We conclude that from a computational point of view it is more attractive to eliminate x_2.

It is important to note that different choices of variables to eliminate may lead to different optimization problems.

- If the constraint contains analysis variables one could make these variables adjustable and use decision rules, thereby introducing much more flexibility. One

can easily prove that when the coefficients for such variables in the equality constraint do not contain uncertain parameters and the equality constraint is linear in the uncertain parameters, then using linear decision rules for such variables is equivalent to eliminating these variables. To be more precise: suppose the linear equality constraint is

$$\mathbf{q}(\boldsymbol{\zeta})^\top \mathbf{x} + y = r,$$

where $\mathbf{q}(\boldsymbol{\zeta})$ is linear in $\boldsymbol{\zeta}$, and y is an analysis variable (without loss of generality we assume the coefficient for y is 1). Then it can easily be proven that substituting $y = r - \mathbf{q}(\boldsymbol{\zeta})^\top \mathbf{x}$ everywhere in the problem is equivalent to using a linear decision rule for y. To reduce the number of extra variables, it is therefore better to eliminate such variables.

- Consider the following robust constraint:

$$(\mathbf{a} + \mathbf{P}\boldsymbol{\zeta})^\top \mathbf{x} = \mathbf{d} \quad \forall \boldsymbol{\zeta} \in Z. \tag{2.17}$$

The equality constraint is satisfied for all $\boldsymbol{\zeta}$ in $\mathcal{Z}$ if $\mathbf{P}^\top \mathbf{x} = \mathbf{0}$. Hence, we could replace (2.17) by the stricter set of equations

$$\mathbf{a}^\top \mathbf{x} = \mathbf{d}, \quad \mathbf{P}^\top \mathbf{x} = \mathbf{0}.$$

However, especially when L is large, this is much too restrictive.

- One could also drop the requirement that the constraints are hard, and make such constraints "soft", by adding, e.g., a quadratic penalty for the violations to the objective.

2.9 On Maximin and Minimax Formulations of RC

In this section, we consider an uncertain LP of the following general form:

$$\max_{\mathbf{x} \geq \mathbf{0}} \{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{d}\},$$

where without loss of generality $\mathbf{A}$ is the uncertain coefficient matrix that resides in the uncertainty set $\mathcal{U}$. So the general RC is given by

$$\text{(R-LP)} \quad \max_{\mathbf{x} \geq \mathbf{0}} \{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{d} \ \ \forall \mathbf{A} \in \mathcal{U}\}.$$

Here we show that (R-LP) can be reformulated as:

$$\text{(RF)} \ \min_{\mathbf{A}\in\mathcal{U}} \max_{\mathbf{x}\geq\mathbf{0}} \{\mathbf{c}^{\top}\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{d}\},$$

if the uncertainty is constraint-wise; however if this condition is not met, then (RF) may not be equivalent to (R-LP).

Remark 1 *This shows that the statement "RO optimizes for the worst case* $\mathbf{A}$*" is too vague. Also the maximin reformulation:*

$$\max_{\mathbf{x}\geq\mathbf{0}} \min_{\mathbf{A}\in\mathcal{U}} \{\mathbf{c}^{\top}\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{d}\},$$

is usually not equivalent to (R-LP). This is because we can almost always find an $\mathbf{x} \geq 0$ *such that no* $\mathbf{A} \in \mathcal{U}$ *exists for which* $\mathbf{A}\mathbf{x} \leq \mathbf{d}$*; therefore, we minimize over an empty set, and have* $+\infty$ *for the maximin objective. Also when* $\mathbf{x}$ *is selected such that at least one feasible* $\mathbf{A}$ *exists (e.g., see Falk (1973)), it is easy to find examples where both formulations are not equivalent.*

To show (R-LP)=(RF) when the uncertainty is constraint-wise, we first take the dual of the (inside) maximization problem of (RF) $[\max_{\mathbf{x}\geq\mathbf{0}} \mathbf{c}^{\top}\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{d}]$. Then, substituting the dual with the primal (maximization) problem in (RF) gives:

$$\text{(OC-LP)} \ \min_{\mathbf{A}\in\mathcal{U},\mathbf{y}\geq\mathbf{0}} \{\mathbf{d}^{\top}\mathbf{y} : \mathbf{A}^{\mathbf{T}}\mathbf{y} \geq \mathbf{c}\},$$

where val(RF)=val(OC-LP) at optimality. Note that the constraints of (RF) can be formulated as $[\mathbf{a_i^T}\mathbf{x} \leq d_i, \forall \mathbf{a_i} \in \mathcal{U}_i, i = 1, \ldots, m]$, if the uncertainty is constraint-wise. Beck and Ben-Tal (2009) show that (OC-LP)—which is the *optimistic counterpart of the dual problem*—is equivalent to the general robust counterpart (R-LP) for constraint-wise uncertainty and disjoint $\mathcal{U}_i$'s. However, if (some of) the constraints are dependent in (R-LP), then we may not sustain the associated equivalence. The following example shows such a situation.

Example

Consider the following toy RC example in which the uncertainty is not constraint-wise:

$$\begin{aligned} \text{(RC-Toy)} \ \max_{\mathbf{y}} \ & y_1 + y_2 \\ & \text{s.t. } a_1 y_1 \leq 1, a_2 y_2 \leq 1 \ \ \forall \mathbf{a} \in \mathbb{R}^2 : ||\mathbf{a}||_2 \leq 1, \end{aligned}$$

where two constraints of the problem are dependent on each other via the ellipsoidal uncertainty set $[\mathbf{a} \in \mathbb{R}^2 : ||\mathbf{a}||_2 \leq 1]$. The robust reformulation of the (RC-Toy) is as follows:

$$\begin{aligned} \text{(RF-Toy)} \quad \min_{\mathbf{a}:||\mathbf{a}||_2 \leq 1} \max_{\mathbf{y}} \quad & y_1 + y_2 \\ \text{s.t.} \quad & a_1 y_1 \leq 1, a_2 y_2 \leq 1, \end{aligned}$$

and the optimistic counterpart (OC) of the problem is

$$\begin{aligned} \text{(OC-Toy)} \quad \min_{\mathbf{x} \geq \mathbf{0},\, \mathbf{a}:||\mathbf{a}||_2 \leq 1} \quad & x_1 + x_2 \\ \text{s.t.} \quad & a_1 x_1 = 1, a_2 x_2 = 1. \end{aligned}$$

(RC-Toy) attains an optimal objective value of 2, whereas the (RF-Toy)'s optimal objective value is $2\sqrt{2}$. Therefore, the robust reformulation (RF-Toy) is not equivalent to the general RC problem (RC-Toy) in this situation. However, val(RF-Toy)=val(OC-Toy) from duality.

2.10 Quality of Robust Solution

In this section we describe how to assess the quality with respect to robustness of a solution based on a simulation study. We first identify four focus points for performing a Monte Carlo experiment, and conclude with two statistical tests that can be used to compare two solutions.

Choice of the uncertainty set. For a comparison between different solutions, it is necessary to define an uncertainty set $\mathcal{U}$ that is used for evaluation. This set should reflect the real-life situation. The uncertainty set that is used for optimization may be different than the set for evaluation. For example, an ellipsoidal set may be used to reduce the conservatism when the real-life uncertainty is a box, while still maintaining a large probability of constraint satisfaction (Ben-Tal et al., 2009, p. 34).

Choice of the probability distribution. A simulation requires knowledge of the probability distribution on the uncertainty set. If this knowledge is ambiguous, it may be necessary to verify whether the simulation results are sensitive with respect to changes in this distribution. For example, Rozenblit (2010) performs different simulations, each based on a probability distribution with a different skewness level.

Choice of the sampling method. For univariate random variables it is computationally easy to draw a random sample from any given distribution. For multivariate

random variables rejection sampling can be used, but it may be inefficient depending on the shape of the uncertainty set, e.g. for an uncertainty set with no volume. A more efficient method for sampling from an arbitrary continuous probability distribution is "hit and run" sampling (Bélisle et al., 1993). An R package for uniform hit and run sampling from a convex body is also available.

Choice of the performance characteristics. From a mathematical point of view there is no difference between uncertainty in the objective and uncertainty in the constraints since an uncertain objective can always be reformulated as a certain objective and an uncertain constraint. However, the distinction between an uncertain objective and an uncertain constraint is important for the interpretation of a solution. First, we look at the effects of adjustable RO and reformulations, then we present the performance characteristics.

Effect of adjustable RO. When one or more "wait and see" variables are modeled as adjustable variables, uncertain parameters may enter the objective function. In that case the performance characteristics for uncertainty in the objective become applicable.

Effect of reformulations. Reformulations are sometimes necessary to end up with a tractable model. The evaluation should be based on the original model, since reformulations introduce additional constraints whose violation is not necessarily a problem. Take for example an inventory model that has constraints on variables that indicate the cost at a certain time period (e.g. constraints (2.18) and (2.19)). These constraints have been introduced to model the costs in the objective function. A violation of these constraints does not render the solution infeasible but does affect the objective value (i.e. the costs of carrying out the solution).

Performance characteristics for uncertainty in the constraints. For an uncertain constraint $f(\mathbf{a}, \boldsymbol{\zeta}) \leq 0$ for all $\boldsymbol{\zeta}$ in $\mathcal{Z}$, the violation is $\max\{0, f(\mathbf{a}, \boldsymbol{\zeta})\}$. Meaningful statistics are the probability on positive violation and the distribution of the violation (average, worst case, standard deviation) under the condition that the violation is positive. When multiple constraints are uncertain, these statistics can be computed per constraint. Additionally, the average number of violated constraints can be reported.

There is a clear trade-off between the objective value and constraint violations. The difference between the worst case objective value of the robust solution and the nom-

inal objective value of the nominal solution is called the *price of robustness* (PoR) (Bertsimas and Sim, 2004). It is useful if the objective is certain, since in that case PoR is the amount that has to be paid for being robust against constraint violations. We observe that PoR is also used when the objective is uncertain. We discourage this, since it compares the nominal solution in case there is no uncertainty with the robust solution where the worst case occurs, so it compares two different scenarios.

Performance characteristics for uncertainty in the objective. Uncertainty in the objective affects the performance of a solution. For every simulated uncertainty vector, the actual objective value can be computed. One may be interested in the worst case, but also in the average value or the standard deviation. For a solution that is carried out many times, reporting the average performance is justified by the law of large numbers. The worst case may be more relevant when a solution is carried out only once or a few times, e.g. when optimizing a medical treatment plan for a single patient. These numbers show what objective value to expect, but they do not provide enough information about the quality of a solution since a high standard deviation is not necessarily undesirable. A robust solution is good when it is close to the *perfect hindsight* (PH) solution. The PH solution is the solution that is obtained by optimizing the decision variables for a specific uncertainty vector as if it is fully known beforehand. This has to be done for every simulated uncertainty vector, and yields an utopia solution. The PH solution may have a large variation, causing a high variation of good solutions as well.

Performance characteristics for any problem. Regardless of whether the uncertainty is in the objective or in the constraints, the mean and associated standard deviation of the difference between the actual performance of a solution and the PH solution are useful for quantifying the quality of a solution. The mean difference between the PH solution and a fully robust solution is defined as the *price of uncertainty* (PoU) by Ben-Tal et al. (2005). It is the maximum amount that a company should invest for reducing the level of uncertainty, e.g. by using more accurate forecasting techniques. It can also be interpreted as the regret of choosing a certain solution rather than the PH solution. Alternative names for PoU are "cost of robustness" (Gregory et al., 2011) or "price of robustness" (Ben-Tal et al., 2004), which are less descriptive than "price of uncertainty" and may cause confusion with price of robustness from (Bertsimas and Sim, 2004). A low mean PoU and a low standard deviation characterize a good solution.

Subtracting the mean objective value of the nominal solution from the mean value

of a robust solution yields the *actual price of robustness* (APoR) (Rozenblit, 2010). APoR can be interpreted as the expected price that has to be paid for using the robust solution rather than the nominal solution, which is negative if RO offers a solution that is better on average. PoR equals APoR when uncertainty only occurs in the constraints.

For multistage problems one may also follow a folding horizon (FH) approach. With FH in each stage where a part of the uncertain parameter is observed, that information is used to optimize for the remaining time periods. This is done by taking the original optimization problem, fixing the decision variables for previous stages, and fixing the elements of the uncertain parameter that have been observed. This allows a fair comparison between a dynamic solution (e.g. created by the AARC) and a static solution (e.g. the nominal solution) when in real-life the static solution is reoptimized in every stage.

Comparing two solutions. We provide several comparison criteria and provide the corresponding statistical test to verify whether one solution is better than another solution. The tests will be demonstrated in Section 2.11. We will assume that the data for the statistics test is available as n pairs (X_i, Y_i) $(i = 1, 2, \ldots, n)$, where X_i and Y_i are performance characteristics in the i'th simulation. For uncertainty in the objective, they can be objective values whereas for uncertainty in the constraints they can be the numbers of constraint violations or the sizes of the constraint violations. We assume that (X_i, Y_i) and (X_j, Y_j) are independent if $i \neq j$, and that smaller values are better. When a conjecture for a test is based on the outcome of a simulation study, the statistical test must be performed with newly generated data to avoid statistical bias. While for the statistical tests it is not necessary that X_i and Y_i are based on the same simulated uncertainty vector $\boldsymbol{\zeta}$, it increases the power of the test since X_i and Y_i will be positively correlated. This reduces the variance of the difference: $\mathrm{Var}(X_i - Y_i) = \mathrm{Var}(X_i) + \mathrm{Var}(Y_i) - 2\,\mathrm{Cov}(X_i, Y_i)$, which is used in the following tests:

- The sign test for the median validates H_0: $m_x = m_y$ against H_1: $m_x < m_y$ with confidence level α, where m_x and m_y are the medians of the distributions of X_i and Y_i, respectively. This tests the conjecture that the probability that solution X outperforms solution Y is larger than 0.5. Let $n_=$ be the number of observations for which $X_i = Y_i$ and let Z be the number of negative signs of $X_i - Y_i$. Under the null hypothesis, Z follows a binomial distribution with parameters $n - n_=$ and 0.5. That means that the null hypothesis gets rejected if Z is larger than the $(1 - \alpha)$ percentile of the binomial distribution.

- The t-test for the mean validates H_0: $\mu_x = \mu_y$ against H_1: $\mu_x < \mu_y$ with confidence level α, where μ_x and μ_y are the means of the distributions of X_i and Y_i, respectively. This tests the conjecture that solution X outperforms solution Y in long run average behavior. This test assumes that $X_i - Y_i$ follows a normal distribution. Let $Z_i = X_i - Y_i$, $\bar{Z} = \sum_{i=1}^{n} Z_i/n$ and $s^2 = \sum_{i=1}^{n}(Z_i - \bar{Z})^2/(n-1)$, then $T = \sqrt{n}\sum_{i=1}^{n}(Z_i - \bar{Z})/s$ follows a t-distribution with $n-1$ degrees of freedom under the null hypothesis. This means that H_0 gets rejected if T is smaller than the α percentile of the t-distribution with $n-1$ degrees of freedom.

2.11 RC May Take Better "Here and Now" Decisions than AARC

A linear decision rule is a linear approximation of a more complicated decision rule. It dictates what to do at each stage as a linear function of observed uncertain parameters, but it is not guaranteed to be the optimal strategy. Every time a decision has to be made it is possible to either follow the linear decision rule, or to reoptimize the AARC for the remaining time periods based on everything that is observed up till then. We will refer to the latter as the AARC-FH, where FH stands for folding horizon. Ben-Tal et al. (2005) compare the AARC with the AARC-FH, and show that the latter produces better solutions on average. A comparison that involves AARC-FH assumes that there is time to reoptimize. It is therefore natural to also make a comparison with the RC-FH, where the RC is solved for the full time horizon and re-optimized for the remaining time period every time a part of the uncertain parameters is unveiled. On average, the RC-FH may outperform the AARC (Cohen et al., 2007; Rozenblit, 2010).

In the remainder of this section we will evaluate both the average and the worst case performance of the nominal solution with FH, the RC-FH and the AARC-FH. A comparison between RC-FH and AARC-FH is new, and shows which model takes the best "here and now" decisions.

We first give an example for the worst case performance. Consider a warehouse that transfers one good. The current inventory is $x_0 = 5$, the holding costs per time period are $h = 1$, the backlogging costs per time period are $b = 2$. In the first period, any nonnegative (not necessarily integer) amount can be ordered while in the second period the maximum order quantity is $q_2^{max} = 3$. Let $T = \{1, 2\}$, let q_t be the order quantity in time period t, and let c_t denote the costs associated with time period t. The ending inventory can be returned to the supplier without penalty fee at time

period three. The optimization problem can be formulated as:

$$
\begin{aligned}
\min \quad & \sum_{t\in T} c_t & & \\
\text{s.t.} \quad & c_t \geq (x_0 + \sum_{i=1}^{T} q_i - d_i)h & \forall t \in T & \quad (2.18)\\
& c_t \geq -(x_0 + \sum_{i=1}^{T} q_i - d_i)b & \forall t \in T & \quad (2.19)\\
& q_2 \leq q_2^{max} & & \\
& q_t \in \mathbb{R}_+ & \forall t \in T. &
\end{aligned}
$$

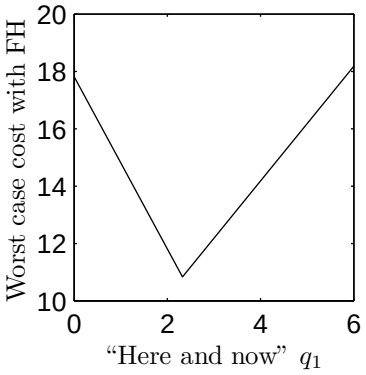
Suppose the demand vector $\mathbf{d}$ is uncertain but is known to reside in a ball around $\mathbf{5}$ with radius 5. We will use this uncertainty set both for optimization and for evaluation.

For this small example, it is possible to approximate the worst case costs for a FH approach as a function of the "here and now" decision q_1 as follows. For each q_1 in a range of values, we have randomly drawn 100 uncertain demand vectors from the boundary of the uncertainty set. For each demand vector we have computed the inventory level at the beginning of the second time period ($= x_0 + q_1 - d_1$). Based on this inventory level, we reoptimized the order quantity for the second time period, where d_2 was assumed to reside in the interval $[5-r, 5+r]$ with $r = \sqrt{25-(d_1-5)^2}$ (so that the full $\mathbf{d}$ vector is in a ball around $\mathbf{5}$ with radius 5). Then we computed the total costs over both time periods. The maximum total costs over all 100 demand vectors approximates the worst case costs with the FH approach, and is depicted in Figure 2.1. From this picture it becomes clear that the optimal order quantity for the first time period is approximately 2.3, which has a worst case performance of 10.8.

We have solved the model for the full time horizon with the RC, with the AARC (where c_1 and c_2 are adjustable on the full $\mathbf{d}$, and q_2 is adjustable on d_1), and as a certain problem with $\mathbf{d} = \mathbf{5}$. The nominal solution gives $q_1 = 0$, the RC gives $q_1 \approx 4.4$, while the AARC yields $q_1 \approx 5.3$, leading to worst case costs of the FH approach of 17.8, 14.9 and 16.8, respectively. So, the RC takes the best "here and now" decision with respect to the worst case performance. It may be paradoxical that the AARC yields a worse solution than the RC, since the feasible region of the AARC includes the RC solution. However, neither of the two optimize the right objective function. Both approximate the objective value using (static or adjustable) auxiliary variables c_t. Wihle AARC indeed has a better objective value than RC, the

solution is not better for the original objective function.

Figure 2.1 – Approximation of the total worst case costs for a FH strategy as a function of the initial order quantity q_1.



We also perform a comparison on a more realistic problem, which is the retailer-supplier flexible commitment problem by Ben-Tal et al. (2005). At the starting time, the retailer commits himself to ordering certain quantities in later months. These commitments are flexible, i.e. deviations are allowed at a penalty cost. The objective is to minimize the total costs for the retailer, consisting of ordering costs (minus the salvage value), holding costs, backlogging costs, penalty costs associated with deviating from the commitments, and costs for variability in the commitments. The article provides two data sets for twelve time periods, A12 and D12, which we also use in our optimization and comparison.

In this problem the retailer faces an uncertain demand. Following Ben-Tal et al. (2005) we consider box uncertainty where the demand may deviate up to $\rho\%$ around the nominal value. For the simulation we draw demand vectors uniformly from this box region. For these demand vectors the nominal solution, RC and AARC are carried out completely. For the FH approach, the reoptimization is performed after each time period based on previously observed demand. 500 demand vectors were generated for each data set and each uncertainty level ρ, and the same demand vectors were used for all models. In addition, the PH solution was computed for each of these demand vectors.

The simulation results are listed in Tables 2.4 and 2.5. For data set A12, the nominal solution with FH results in the lowest average costs. This means that the nominal solution takes better "here and now" decisions than RC and AARC. Moreover, the RC-FH has lower average costs than the AARC-FH, so also the RC takes better "here and now" decisions than AARC. The advantage of the nominal FH solution

compared to RC-FH and AARC-FH increases when the uncertainty set becomes larger. For data set W12 the nominal solution is the best solution and FH leads to higher mean costs. For this data set, AARC takes significantly better "here and now" decisions than RC. When comparing Nominal-FH with AARC-FH in Table 2.4 it is not immediately clear which solution is better, since the lower mean value of Nominal-FH comes with a larger standard deviation. The Nominal-FH in Table 2.4 is an example of the statement in Section 2.10 that a high standard deviation is not necessarily bad. Its standard deviation is higher than that of AARC-FH but this is due to the high standard deviation of PH, as can be seen from Table 2.5. From this last table, it can be seen that Nominal-FH is strictly better than AARC-FH. For data set A12 and $\rho = 10\%$, it is not clear whether RC-FH outperforms AARC-FH. We now demonstrate the two statistical tests from Section 2.10 on this data set, each based on 100 newly generated uncertainty vectors, to test whether the RC-FH outperforms the AARC-FH. The null hypothesis that both solutions perform equal on average is rejected ($p = 6.1 \cdot 10^{-6}$), and also the null hypothesis that the medians of RC-FH and AARC-FH are equal is rejected ($p = 1.4 \cdot 10^{-10}$). These results show that the AARC is not necessarily better than the RC and support the statement in Section 2.10 that a simulation is required for comparing solutions. As mentioned in Section 2.5, RO may provide solutions that are not Pareto efficient when multiple optimal solutions exist. A different optimal solution to the RC or AARC may yield completely different simulation results, rendering our conclusions useless. This is not the case. We have verified this by solving the problem in two stages. In the first stage we solve the robust counterpart (RC or AARC). For the second stage we take the same problem as in the first stage, but we add the constraint that the robust objective value is not worse than the optimal value from the first stage, and we change the objective in minimizing the costs for the nominal demand trajectory. Thus, we find a solution that is robust optimal and that cannot be improved with respect to the nominal demand trajectory. Moreover, the resulting solution is Pareto optimal; see Iancu and Trichakis (2013). The second stage problem returns the same solution as the first stage problem, so our conclusions are unaffected.

Table 2.4 – Simulated mean (std) costs for the retailer-supplier flexible commitment problem.

	A12		W12	
	$\rho = 10\%$	$\rho = 50\%$	$\rho = 10\%$	$\rho = 50\%$
Nominal	688 (35)	848 (211)	12775 (708)	16163 (3899)
RC	731 (14)	1140 (70)	13656 (169)	20046 (1251)
AARC	702 (5)	1071 (53)	13314 (35)	18575 (192)
Nominal-FH	674 (14)	774 (87)	12869 (296)	16280 (1251)
RC-FH	699 (5)	979 (19)	13615 (125)	19260 (585)
AARC-FH	700 (5)	1027 (21)	13314 (35)	18572 (192)
PH	658 (11)	699 (50)	12194 (204)	12911 (1144)

Table 2.5 – Simulated mean (std) PoU for the retailer-supplier flexible commitment problem.

	A12		W12	
	$\rho = 10\%$	$\rho = 50\%$	$\rho = 10\%$	$\rho = 50\%$
Nominal	30 (29)	149 (178)	581 (575)	3252 (3186)
RC	73 (19)	441 (89)	1463 (308)	7135 (2266)
AARC	44 (9)	372 (74)	1120 (230)	5664 (1298)
Nominal-FH	16 (6)	75 (44)	675 (144)	3369 (717)
RC-FH	41 (9)	280 (49)	1421 (253)	6349 (1550)
AARC-FH	42 (9)	328 (51)	1120 (230)	5661 (1301)

2.12 Conclusion

In this chapter, we have presented a general recipe that shall be helpful for using RO in practice. Additionally, we give several practical insights and hints in applying RO. Examples of such practical insights are: the robust reformulations of equivalent deterministic optimization problems may not be equivalent; in multi-stage optimization problems, re-optimizing the given problem at each stage using static RO or nominal data may outperform solutions provided by ARO; and the actual probability guarantee of an uncertainty set is often higher than the probabilistic guarantee that is approximated by using a safe approximation technique. We also discuss many practical issues to apply RO in a successful and convincing way. Examples are: how to choose the uncertainty set; what is the right interpretation of "RO optimizes for the worst case"; and should the decision rule used in ARO be a function of the final or the primitive uncertainty? Moreover, we propose ideas on how to deal with equality constraints and integer adjustable variables, and on how to compare the robustness characteristics of two solutions. We have provided many numerical examples to illustrate our insights and discussions, and to demonstrate the effectiveness of the usefulness of our hints.

Bibliography

Bandi, C. and Bertsimas, D. (2012). Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical Programming*, 134(1):23–70.

Beck, A. and Ben-Tal, A. (2009). Duality in robust optimization: primal worst equals dual best. *Operations Research Letters*, 37(1):1–6.

Bélisle, C. J. P., Romeijn, H. E., and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266.

Ben-Tal, A., den Hertog, D., De Waegenaere, A. M. B., Melenberg, B., and Rennen, G. R. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

Ben-Tal, A., den Hertog, D., and Vial, J.-P. (2012). Deriving robust counterparts of nonlinear uncertain inequalities. *CentER Discussion Paper*, 2012(053).

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press.

Ben-Tal, A., Golany, B., Nemirovski, A., and Vial, J.-P. (2005). Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management*, 7(3):248–271.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

Ben-Tal, A. and Nemirovski, A. (2002). Robust optimization–methodology and applications. *Mathematical Programming*, 92(3):453–480.

Ben-Tal, A. and Nemirovski, A. (2008). Selected topics in robust convex optimization. *Mathematical Programming*, 112(1):125–158.

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501.

Bertsimas, D. and Caramanis, C. (2010). Finite adaptability in multistage linear

optimization. *IEEE Transactions on Automatic Control*, 55(12):2751–2766.

Bertsimas, D. and Georghiou, A. (2013). Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Optimization Online*, Sept.

Bertsimas, D. and Sim, M. (2004). The price of robustness. *Operations Research*, 52(1):35–53.

Bertsimas, D. and Thiele, A. (2004). A robust optimization approach to supply chain management. In *Integer programming and combinatorial optimization*, pages 86–100.

Bienstock, D. and Özbay, N. (2008). Computing robust basestock levels. *Discrete Optimization*, 5(2):389–414.

Birge, J. R. and Louveaux, F. V. (2011). *Introduction to Stochastic Programming*. Springer.

Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research logistics quarterly*, 9(3-4):181–186.

Chen, X., Sim, M., Sun, P., and Zhang, J. (2008). A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357.

Cohen, I., Golany, B., and Shtub, A. (2007). The stochastic time-cost tradeoff problem: A robust optimization approach. *Networks*, 49(2):175–188.

de Ruiter, F. (2013). Adjustable robust optimization based on inexact information. Master's thesis, Tilburg University, the Netherlands.

Falk, J. E. (1973). A linear max–min problem. *Mathematical Programming*, 5(1):169–188.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fredriksson, A., Forsgren, A., and Hårdemark, B. (2011). Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys*, 38(3):1672–1684.

Gorissen, B. L. (2013). Robust fractional programming. Working paper.

Gorissen, B. L., Ben-Tal, A., Blanc, H., and den Hertog, D. (2012). A new method for deriving robust and globalized robust solutions of uncertain linear conic optimization problems having general convex uncertainty sets. *CentER Discussion Paper*, 2012(076).

Gorissen, B. L. and den Hertog, D. (2012). Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–43.

Gregory, C., Darby-Dowman, K., and Mitra, G. (2011). Robust optimization and portfolio selection: The cost of robustness. *European Journal of Operational Re-*

search, 212(2):417–428.

Hadjiyiannis, M. J., Goulart, P. J., and Kuhn, D. (2011). A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 7386–7391. IEEE.

Iancu, D. and Trichakis, N. (2013). Pareto efficiency in robust optimization. *Management Science*, online first.

Kuhn, D., Wiesemann, W., and Georghiou, A. (2011). Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209.

Lobo, M. S. (2000). *Robust and convex optimization with applications in finance*. PhD thesis, Stanford University.

Lobo, M. S., Vandenberghe, L., Boyd, S., and Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1):193–228.

Pardo, L. (2005). *Statistical inference based on divergence measures*. CRC Press.

Postek, K. (2013). Adjustable robust optimization by splitting the uncertainty set. Master's thesis, Tilburg University, the Netherlands.

Prekopa, A. (1995). *Stochastic Programming*. Kluwer Academic Publishers.

Roelofs, M. and Bisschop, J. (2012). AIMMS 3.12: The Language Reference, Paragon Decision Technology BV.

Rozenblit, M. (2010). Robust optimization of a multi-period production planning problem under uncertainty. Master's thesis, Technion – Israel Institute of Technology.

Shapiro, A. and Ruszczyński, A. P. (2003). *Stochastic Programming*. Elsevier.

Vayanos, P., Kuhn, D., and Rustem, B. (2011). Decision rules for information discovery in multi-stage stochastic programming. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 7368–7373. IEEE.

CHAPTER 3

Safe Approximations of Ambiguous Chance Constraints Using Historical Data

3.1 Introduction

The objective of *robust optimization* (RO) is to find solutions that are immune to the uncertainty of the parameters in a mathematical optimization problem. It requires that the constraints of a given problem should be satisfied for all realizations of the uncertain parameters in a so-called uncertainty set. The robust version of a mathematical optimization problem is generally referred to as the *robust counterpart* (RC) problem. RO is popular because of the tractability of the RC for many classes of uncertainty sets. For example, the RC of an uncertain linear optimization problem with data varying in a polyhedral uncertainty set can be reformulated as a linear optimization (LO) problem (Ben-Tal et al. 2009, Ben-Tal and Nemirovski 2002). Additionally, the RC of an uncertain LO problem with an ellipsoidal uncertainty set can be reformulated as a second-order cone problem (SOCP) that can be solved efficiently by existing solvers. The choice of the uncertainty set is important for two reasons. First, it plays a critical role in the tractability of the RC problem. Second, it should represent the actual uncertainty in a meaningful way.

The main criticism for RO is that it finds a feasible solution for all the uncertainty—supported in the so-called uncertainty set—regardless of the occurrence probabilities of the uncertainties, and the price of robustness can be paid as an overconservative solution. A remedy for this criticism resides in defining the uncertainty set using a safe approximation of the chance constraint. Later in this section we dwell into the details of the safe approximation methods, but before that we briefly discuss chance constrained optimization that is introduced by Charnes et al. (1958), Charnes and Cooper (1959), Miller and Wagner (1965) and Prékopa (1970). A *chance constraint* is given by

$$\Pr_{\zeta \sim \mathbb{P}} \{\zeta : f(\mathbf{x}, \zeta) \leq 0\} \geq \beta, \tag{3.1}$$

where $f(\mathbf{x}, \zeta)$ denotes a function of a decision vector $\mathbf{x} \in \mathbb{R}^n$ and a random perturbation vector $\zeta \in \mathbb{R}^\ell$, $\beta \in [0, 1]$ is the prescribed probability bound and $\mathbb{P}$ is the known probability distribution of ζ. Different than RO, in chance constrained optimization it is assumed that the uncertain parameters are distributed according to a known distribution. It can be shown that the feasible set of (3.1) is convex when $f(x, \zeta) = g(x) - \zeta$ and ζ follows a log-concave distribution (Prékopa 1973, Prékopa 1995). Additional convexity results for (3.1) can be shown when ζ follows a Gaussian distribution; see van de Panne and Popp (1963), Prékopa (1974), Burkauskas (1986) and Henrion and Strugarek (2008). It should be stressed that even though chance constraints are tractable for the cases that are mentioned above, in general they are computationally intractable. The reason is twofold. First, the feasible set of the chance constraint is usually non-convex. Second, even if the feasible set is convex, it can be intractable to compute the left-hand side of the constraint (e.g., typical way of checking feasibility is via Monte Carlo simulation which is very costly for high accuracies). Moreover, the probability distribution $\mathbb{P}$ that has to be known to compute (3.1), is often not (exactly) known in practice.

Computationally tractable safe approximations of the chance constraint have recently been proposed to overcome the difficulties that are mentioned above. A set of constraints is called a safe approximation of the chance constraint if the feasible set of the safe approximation is a subset of the feasible set of the chance constraint. The seminal work of Nemirovski and Shapiro (2006) is based on building a computationally tractable approximation of a chance constrained problem. The authors assume that the constraints are affine and entries of the perturbation vector, so-called uncertain parameters, are independent with known support. Ben-Tal and Nemirovski (2000) propose safe convex approximations of scalar chance constraints. The authors translate the existing stochastic uncertainties to "uncertain-but-bounded" sets assuming that the uncertain parameters are mutually independent with zero mean. The obtained approximations in Nemirovski and Shapiro (2006) and Ben-Tal and Nemirovski (2000) are computationally tractable and perform good when the number of uncertain parameters is relatively high. In addition, Ben-Tal and Nemirovski (2009) elaborate a safe tractable approximation of the chance constrained version of an affinely perturbed linear matrix inequality (LMI) constraint, assuming that the primitive uncertain parameters are independent with light-tail distributions (e.g., bounded or Gaussian). Chen et al. (2010) propose an alternative conservative approximation of a joint chance constraint in terms of a worst-case conditional value-at-risk (CVaR) constraint. The resulting approximation outperforms the Bonferroni approximation. Zymler et al. (2011) develop new tools and models for approxi-

mating joint chance constraints under the assumption that first- and second-order moments together with the support of the perturbation vector are known. The authors propose an efficient sequential semidefinite programming (SDP) algorithm to solve distributionally robust chance constraint program. As an alternative to the safe approximation techniques that are mentioned above, Calafiore and Campi (2005) and Campi and Garatti (2008) propose a 'randomized' approach. The randomized approach substitutes the original semi-infinite uncertain constraint with a finite set of constraints which are randomly sampled from the original constraint according to the known distribution of the uncertain parameter. The authors show that the resulting randomized solution fails to satisfy only a small proportion of the original constraints provided that a sufficient number of samples is drawn.

It is important to point out that in practice we usually have partial or no information on the probability distribution $\mathbb{P}$, since it needs to be estimated from historical data. This is why it makes sense to pass to *ambiguous* chance constraint. The term "ambiguous" stands for the uncertainty in the probability distribution. In other words, the distribution of the uncertain parameters is itself uncertain. Different than the classical chance constraint, in the ambiguous case $\mathbb{P}$ belongs to a family of distributions $\mathcal{P}$ and the chance constraint is satisfied for all probability distributions in $\mathcal{P}$, i.e.,

$$\Pr_{\zeta\sim\mathbb{P}}\{\zeta : f(\mathbf{x},\zeta)\leq 0\}\geq\beta, \quad \forall\mathbb{P}\in\mathcal{P}. \tag{3.2}$$

This introduces an additional complexity in solving the problem aside from the existing difficulties that are mentioned above. Formulating ambiguity in the probability distribution has taken attention of scholars from different fields. A wide variety optimization problems under uncertainty involves an expectation function such as $[\min_{x\in X}\mathbb{E}_{\mathbb{P}}[f(x,\zeta)]]$, where $X\subseteq\mathbb{R}^n$ is the set of feasible decisions and $\mathbb{P}$ is the known probability distribution of ζ. In the absence of full information on the probability distribution or in other words when only a set of possible distributions $\mathcal{P}$ is known, it is natural to optimize the expectation function corresponding to the worst-case probability distribution in $\mathcal{P}$. This lead to the following *minimax* formulation: $[\min_{x\in X}\{\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[f(x,\zeta)]\}]$. For more details of such formulations, we refer to Žáčková (1966), Birge and Wets (1986), Dupačová (1987), Dupačová (2001), Shapiro and Kleywegt (2002) and Shapiro and Ahmed (2004). Moreover, in economics, ambiguity in the probability distribution is addressed by Gilboa and Schmeidler (1989), Hansen and Sargent (2001) and Epstein and Schneider (2003, 2007).

Related to this chapter, ambiguity in the context of the chance constrained optimization has recently been studied by Erdoğan and Iyengar (2006) and Nemirovski and

Shapiro (2007). In the seminal work of Erdoğan and Iyengar (2006), the distribution family $\mathcal{P}$ in (3.2) is defined using the Prohorov metric. The authors propose a robust sampled problem that is a good approximation for the associated ambiguous chance constrained problem with high probability. Nemirovski and Shapiro (2007) show that the convex approximation of the regular chance constraint that is proposed in the paper straightforwardly extends to the ambiguous case when $\mathcal{P}$ is known to belong to a given convex compact set. As it is stated by Ben-Tal et al. (2009, p. 30), the definition of a safe tractable approximation of the chance constraint straightforwardly extends to the ambiguous chance constraint and for this reason the adjective "ambiguous" is generally skipped. Similarly, for the ease of exposition in the sequel, we usually skip the adjective "ambiguous" and the general representation in (3.2). However, it is important to point out that the ambiguous chance constraint is "severely" intractable compared to the regular chance constraint, even though its extension for the safe approximations is straightforward and tractable.

In this chapter, we propose a safe approximation of ambiguous joint chance constraints where the family of distributions $\mathcal{P}$ in (3.2) is given by a confidence set that is based on the historical data information and goodness-of-fit statistics known in the statistics literature as ϕ-divergence. Numerical results show that it leads to tighter uncertainty sets compared to the existing safe approximation methods, and therefore yields better objective values for the uncertain problem under consideration. The new approach is suitable for dependent and independent uncertain parameters, and can be extended to nonlinear inequalities. Most importantly, the proposed methodology does not require the assumption that the probability distribution or the certain moments of the uncertain parameters are known. The disadvantage of our approach is that it requires extensive data when the number of uncertain parameters is high.

Research that is related to ϕ-divergence includes the following. Klabjan et al. (2007) and Calafiore (2007) use two special cases of ϕ-divergence to construct uncertainty regions from historical data. The former derives the robust stochastic lot-sizing problem and uses χ^2-statistics; the latter formulates the robust portfolio selection problem and considers Kullback-Leibler divergence. In both papers, the uncertain parameters are probability vectors, and the goal is to find robust solutions that are feasible for all allowable distributions of the uncertain parameters with bounded support. Ben-Tal et al. (2012) take up the topic under the more general title of ϕ-divergence and focus on robust optimization problems with uncertainty regions defined by ϕ-divergence distance measures. They provide tractable formulations of robust optimization problems for ϕ-divergence-based uncertainty regions. Their results show that uncertainty

sets based on ϕ-divergence are good alternatives for the uncertainty sets such as ellipsoidal, box, and their variations that are well studied in the literature. In this chapter, we go one step further and use ϕ-divergence-based uncertainty sets not only for uncertain probability vectors but also for general uncertain parameters.

The remainder of the chapter is organized as follows. In §3.2, we give an introduction to ϕ-divergence and confidence sets. In §3.3, we discuss the new safe approximation method. Then, in §3.4, we present the results of several numerical experiments. Finally, we provide concluding remarks in §3.5.

3.2 Introduction to ϕ-Divergence and Confidence Sets for Probability Vectors

In this section we define ϕ-divergence and some of the properties taken from Ben-Tal et al. (2012), Pardo (2006) and Jager and Weller (2007) that are used in later sections.

3.2.1 Confidence Sets Based on ϕ-Divergence

Given N historical observations on an uncertain parameter $\zeta \in \mathbb{R}^{\ell}$, the support of ζ is divided into m cells such that the number of observations o_i in cell $i \in \{1, \ldots, m\}$ is at least five:

$$\sum_{i=1}^{m} o_i = N \text{ such that } o_i \geq 5 \quad \forall i \in \{1, \ldots, m\}.$$

Then, the historical data on ζ are translated into frequencies $q = (q_1, ..., q_m)$ such that $e^T q = 1$, where e is the all-one vector and q_i is the observed frequency of cell $i \in \{1, \ldots, m\}$ given by

$$q_i = \frac{o_i}{N}.$$

We consider p_i to be the unknown true probability vector for cell i, i.e., $p_i = \Pr(\zeta \in \text{cell i})$, and we construct a confidence set for p using the empirical estimate q and goodness-of-fit statistics, more precisely ϕ-divergence. The ϕ-divergence ("distance") between two vectors $p = (p_1, ..., p_m) \geq 0$ and $q = (q_1, ..., q_m) \geq 0$ in $\mathbb{R}^m$ is defined by

$$I_{\phi}(p, q) := \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right), \tag{3.3}$$

where $\phi(t)$ is convex for $t \geq 0$, $\phi(1) = 0$, $\phi(a/0) := a \lim_{t\to\infty} \phi(t)/t$ for $a > 0$, and $\phi(0/0) = 0$. If we assume that ϕ is twice continuously differentiable in the neighborhood of 1 and $\phi''(1) > 0$, then the test statistic

$$\frac{2N}{\phi''(1)} I_\phi(p, q)$$

asymptotically follows a χ^2_{m-1}-distribution with $(m-1)$ degrees of freedom. Using this test statistic, an approximate $(1-\alpha)$-confidence set for p defines the family of distributions $\mathcal{P}$ in (3.2):

$$\mathcal{P} := \left\{ p \in \mathbb{R}^m : p \geq 0, p^T e = 1, I_\phi(p, q) \leq \rho \right\}, \tag{3.4}$$

where

$$\rho := \frac{\phi''(1)}{2N} \chi^2_{m-1,1-\alpha}. \tag{3.5}$$

Different choices of $\phi(.)$ have been studied in the literature. See Pardo (2006), Jager and Weller (2007), Ben-Tal et al. (2012) and Gushchin (2008) for an overview; Table 4.1 taken from Ben-Tal et al. (2012) presents the most common choices of $\phi(.)$ together with the conjugate function that is defined as follows:

$$\phi^*(s) := \sup_{t \geq 0} \{st - \phi(t)\}.$$

In this chapter, we work with ϕ-divergence distances for which the closed-form conjugates are available; see Table 4.1.

Table 3.1 – ϕ-Divergence Examples

Divergence	$\phi(t), t > 0$	$I_\phi(p, q)$	$\phi^*(s)$
Kullback-Leibler	$t \log t$	$\sum_i p_i \log\left(\frac{p_i}{q_i}\right)$	e^{s-1}
Burg entropy	$-\log t$	$\sum_i q_i \log\left(\frac{p_i}{q_i}\right)$	$-1 - \log(-s), s \leq 0$
χ^2-distance	$\frac{1}{t}(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{p_i}$	$2 - 2\sqrt{1-s}, s \leq 1$
Pearson χ^2-distance	$(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{q_i}$	$s + s^2/4, s \geq -2$ $-1, s < -2$
Hellinger distance	$(1-\sqrt{t})^2$	$\sum_i (\sqrt{p_i} - \sqrt{q_i})^2$	$\frac{s}{1-s}, s \leq 1$

3.2.2 Probability Bound for Subset of Cells

Let $V = \{1, \ldots, m\}$ be the set of cell's indices. The uncertainty region that is formed by the union of cells in $S \subseteq V$ is given by

$$C(S) = \bigcup_{i \in S} (\text{cell i}). \tag{3.6}$$

In our approach, we choose S such that $\Pr_\zeta(\zeta \in C(S)) \geq \beta$, where ζ is the primitive uncertain parameter and β is the prescribed probability bound in the chance constraint (3.2). The structure of the cells and how to find S will be clarified in §3.3.1 and §3.3.2, respectively. In this subsection, we determine a probability guarantee for a given S. To do this we calculate the minimal value of $\sum_{i \in S} p_i$ such that p is in the $(1-\alpha)$-confidence set (3.4):

$$\begin{aligned}
(\text{P}) \quad \gamma(S, \alpha) = \min \ & \sum_{i \in S} p_i && (3.7)\\
\text{s.t.} \ & I_\phi(p, q) \leq \frac{\phi''(1)}{2N} \chi^2_{m-1,1-\alpha} (= \rho) && (3.8)\\
& \sum_{i \in V} p_i = 1 && (3.9)\\
& p \geq 0. && (3.10)
\end{aligned}$$

Note that (P) is a convex optimization problem in $p \in \mathbb{R}^{|V|}$ since ϕ-divergence functions are convex. Constraints (3.8) to (3.10) define a $(1-\alpha)$-confidence set, and the probability that the uncertain parameter is in the region defined by S, is at least $\gamma(S, \alpha)$ with a $(1-\alpha)$ confidence level.

The following theorem shows an alternative way of calculating $\gamma(S, \alpha)$ by using the dual problem of (P).

Theorem 1 *Suppose $\phi(.)$ is convex and $\alpha < 1$, then the optimal objective value of problem (P) is equal to the optimal objective value of the following lagrangian dual (LD) problem:*

$$(LD) \quad \max_{\eta \geq 0, \, \lambda} \left\{ -\eta\rho - \lambda - \eta \left[\phi^* \left(-\frac{\lambda + 1}{\eta} \right) \sum_{i \in S} q_i + \phi^* \left(-\frac{\lambda}{\eta} \right) \sum_{i \in V \setminus S} q_i \right] \right\}$$

in which $\phi^*(s) = \sup_{t \geq 0} \{st - \phi(t)\}$.

Proof. The objective function of (P) can be rewritten as $\sum_{i \in V} a_i p_i$, where

$$a_i = \begin{cases} 1, & i \in S \\ 0, & \text{otherwise.} \end{cases}$$

Then we can derive the Lagrangian function as follows:

$$\begin{aligned} L(p,\eta,\lambda) &= \sum_{i\in V} a_i p_i + \lambda\left(\sum_{i\in V} p_i - 1\right) + \eta\left(I_\phi(p,q) - \rho\right) \\ &= -\eta\rho - \lambda + \sum_{i\in V}\left[(\lambda + a_i)p_i\right] + \eta I_\phi(p,q). \end{aligned}$$

The corresponding Lagrangian objective function is as follows:

$$\begin{aligned} g(\lambda,\eta) &= \min_{p\geq 0} L(p,\eta,\lambda) \\ &= -\eta\rho - \lambda + \min_{p\geq 0}\left\{\sum_{i\in V}\left[(\lambda + a_i)p_i + \eta q_i \phi\left(\frac{p_i}{q_i}\right)\right]\right\} \\ &= -\eta\rho - \lambda + \min_{p\geq 0}\left\{\sum_{i\in V} -\eta q_i\left[-\frac{(\lambda + a_i)}{\eta}\frac{p_i}{q_i} - \phi\left(\frac{p_i}{q_i}\right)\right]\right\}. \end{aligned}$$

In the last term of the above formulation we have used (3.3). Then the Lagrangian objective is equivalent to the following:

$$\begin{aligned} g(\lambda,\eta) &= -\eta\rho - \lambda - \max_{p\geq 0}\left\{\sum_{i\in V}\eta q_i\left[\left(\frac{-\lambda - a_i}{\eta}\right)\frac{p_i}{q_i} - \phi\left(\frac{p_i}{q_i}\right)\right]\right\} \\ &= -\eta\rho - \lambda - \sum_{i\in V}\eta q_i \max_{p\geq 0}\left\{\left[\left(\frac{-\lambda - a_i}{\eta}\right)\frac{p_i}{q_i} - \phi\left(\frac{p_i}{q_i}\right)\right]\right\} \\ &= -\eta\rho - \lambda - \eta\sum_{i\in V}\left[q_i\phi^*\left(-\frac{\lambda + a_i}{\eta}\right)\right] \\ &= -\eta\rho - \lambda - \eta\left[\sum_{i\in S} q_i\phi^*\left(-\frac{\lambda + 1}{\eta}\right) + \sum_{i\in V\setminus S} q_i\phi^*\left(-\frac{\lambda}{\eta}\right)\right], \end{aligned}$$

where

$$\phi^*(s) := \sup_{t\geq 0}\left\{st - \phi(t)\right\}.$$

Finally, the Lagrangian Dual Problem is the maximization problem presented below:

$$\begin{aligned} &(\mathrm{LD}) \max_{\eta\geq 0,\,\lambda}\left\{g(\lambda,\eta)\right\} \\ &\quad = \max_{\eta\geq 0,\,\lambda}\left\{-\eta\rho - \lambda - \eta\left[\sum_{i\in S} q_i\phi^*\left(-\frac{\lambda + 1}{\eta}\right) + \sum_{i\in V\setminus S} q_i\phi^*\left(-\frac{\lambda}{\eta}\right)\right]\right\}. \end{aligned}$$

■

The dual problem is an optimization problem with two variables (η, λ) and a simple constraint $\eta \geq 0$. Furthermore, the convexity of $\phi^*(\lambda)$ implies that $\eta\phi^*\left(\frac{\lambda}{\eta}\right)$ is jointly convex in λ and η. Hence, (LD) is a convex optimization problem with only two variables and it can be solved efficiently.

Independent uncertain parameters. In some cases it may be known that the uncertain parameters ζ_j are independent for $j \in \{1, \ldots, \ell\}$. Let V_j denote the set of cells for the j^{th} uncertain parameter and $m_j = |V_j|$. Since the uncertain parameters are independent, we may have historical data for each uncertain parameter separately, and N_j denotes the sample size of the data for the j^{th} parameter. In addition, the probability that the j^{th} uncertain parameter is in cell $i \in V_j$ is denoted by $p_i^{(j)}$. Similarly, $q_i^{(j)}$ denotes the frequency of cell $i \in V_j$ for the j^{th} uncertain parameter. An aggregate cell is indexed by $(i_1, i_2 \ldots, i_\ell)$, where $i_j \in V_j$ for all $j \in \{1, \ldots, \ell\}$. Because of the independence, the probability that the uncertain parameters are in cell $(i_1, i_2 \ldots, i_\ell)$ is equivalent to

$$p_{i_1,i_2\ldots,i_\ell} = \prod_{j=1}^{\ell} p_{i_j}^{(j)},$$

and the frequency of cell $(i_1, i_2 \ldots, i_\ell)$ is given by

$$q_{i_1,i_2\ldots,i_\ell} = \prod_{j=1}^{\ell} q_{i_j}^{(j)}.$$

All elements $p_{i_1,i_2\ldots,i_\ell}$ are collected in vector $p \in \mathbb{R}^{m_1 m_2 \ldots m_\ell}$, according to the order of the indices. Similarly, $q_{i_1,i_2\ldots,i_\ell}$ are collected in vector $q \in \mathbb{R}^{m_1 m_2 \ldots m_\ell}$. Then, the following mathematical optimization problem is a special case of (P), in the case of ℓ independent parameters:

$$\begin{aligned}
\text{(IP)} \quad \min \quad & \sum_{(i_1,i_2,\ldots,i_\ell)\in S} p_{i_1,i_2\ldots,i_\ell} \\
\text{s.t.} \quad & I_\phi(p,q) \leq \frac{\phi''(1)}{2N_1N_2\ldots N_\ell}\chi^2_{(m_1-1)(m_2-1)\ldots(m_\ell-1),1-\alpha} && (3.11)\\
& \sum_{i\in V_j} p_i^{(j)} = 1 \quad \forall j \in \{1,\ldots,\ell\} && (3.12)\\
& p_{i_1,i_2\ldots,i_\ell} = \prod_{j=1}^{\ell} p_{i_j}^{(j)} \quad \forall i_j \in V_j, \forall j \in \{1,\ldots,\ell\} && (3.13)\\
& p_i^{(j)} \geq 0 \quad \forall i \in V_j, \forall j \in \{1,\ldots,\ell\}, && (3.14)
\end{aligned}$$

where $(m_1 - 1)(m_2 - 1)\ldots(m_\ell - 1)$ denotes the degrees of freedom when we have ℓ independent parameters. It is easy to see that (IP) has highly nonlinear terms in

constraint (3.13) and is nonconvex. Fortunately, the following theorem relaxes the nonlinear structure of (IP) and provides a lower bound for the objective function.

Theorem 2 *Let* $V = V_1 \times V_2 ... \times V_\ell$, $m - 1 = (m_1 - 1)(m_2 - 1)\dots(m_\ell - 1)$, $N = N_1 N_2 \dots N_\ell$, *and* $S \subseteq V$, *then (P) is a relaxation of (IP).*

Proof. Let $(\hat{p}, \hat{p}^{(1)}, \dots, \hat{p}^{(\ell)})$ be a feasible solution of (IP). If we prove that $\hat{p} \in \mathbb{R}^{m_1 m_2 \cdots m_\ell}$ of $(\hat{p}, \hat{p}^{(1)}, \dots, \hat{p}^{(\ell)})$ are feasible for (P), then we can conclude that (P) is a reduced relaxation of (IP).
To begin with, let $V = V_1 \times V_2 ... \times V_\ell$, $m - 1 = (m_1 - 1)(m_2 - 1)\dots(m_\ell - 1)$ and $N = N_1 N_2 \dots N_\ell$. Then, constraint (3.11) in (IP) coincides with constraint (3.8) in (P). In addition, constraints (3.12) and (3.13) imply that the $\hat{p}$ values sum up to 1. Moreover, from constraints (3.13) and (3.14) in (IP), it is easy to verify that $\hat{p}_{i_1,i_2,\dots,i_\ell} \geq 0$ for all $i = (i_1, \dots, i_\ell) \in V$. As a result, $\hat{p}$ satisfy all the constraints in (P). ■

Note that the optimal solution $\hat{p}$ of (P) does not necessarily satisfy (3.13) for the individual probabilities $\hat{p}_{i_j}^{(j)}$ given by $[\hat{p}_{i_j}^{(j)} = \sum_{k \neq j}^{\ell} \sum_{i_k \in V_k} \hat{p}_{i_1,.,i_j,.,i_\ell}]$, and hence the elements of $\hat{p}$ may not be independent. However, we are looking for a good lower-bound probability for S that can be computed efficiently. This is why we use (P), or equivalently (LD) in Theorem 1, that yields a tight probability bound $\gamma(S, \alpha)$ for any given parameter structure that can be dependent or independent. Nevertheless, working with independent uncertain parameters has some advantages compared to the dependent case. First, we obtain tighter $(1 - \alpha)$-confidence sets for p. This is because we have fewer degrees of freedom for the same number of cells, so the ρ value gets smaller in (3.8). Second, the sample size becomes the product of the individual sample sizes; see Theorem 2, and we require fewer data.

3.3 Safe Approximation Method

In this section, we provide our method to derive safe approximations for chance constrained problems. We first describe the general setup of our approach and then explain the details of each step in our algorithm. Finally, we mention possible extensions of the algorithm to joint chance constraints and nonlinear inequalities.

3.3.1 General Setup

For the sake of simplicity, we explain our safe approximation method for linear optimization. Later in §3.3.3.2, it is shown how the method is extended to nonlinear

inequalities.
We consider the following chance constrained linear optimization problem:

$$\begin{aligned}(\text{ULO}) \quad & \max_{x\in\mathbb{X}} \; c^{\mathrm{T}}x \\ & \text{s.t. } \Pr_{\zeta}\{\zeta \in [-1,1]^{\ell} : a_k(\zeta)^{\mathrm{T}}x \le b_k, \forall k \in \{1,...,K\}\} \ge \beta, \end{aligned} \tag{3.15}$$

where $x \in \mathbb{R}^n$ is a vector of decision variables, $\mathbb{X}$ is the feasible set given by deterministic constraints, $c \in \mathbb{R}^n$ is a vector of objective coefficients, $b \in \mathbb{R}^K$ is a vector of right-hand side values, β is the given probability bound and $a_k(\zeta) \in \mathbb{R}^n$ is linear in the primitive uncertain parameter $\zeta \in [-1,1]^{\ell}$, i.e.,

$$a_k(\zeta) = a_k^0 + \sum_{j=1}^{\ell} \zeta_j a_k^j \quad \forall k \in \{1,...,K\}, \tag{3.16}$$

where $a_k^j \in \mathbb{R}^n$, $j \in \{0,\ldots,\ell\}$. We may assume w.l.o.g. that $\zeta \in [-1,1]^{\ell}$, since scaling for different intervals can be done by adjusting all the a_k^j. We may also assume w.lo.g that the right-hand side vector b is certain, since the uncertain right-hand side can easily be reformulated in RO. Moreover, we assume that the number of uncertain parameters, ℓ, is much smaller than n. This is motivated by the fact that in many cases a few primitive sources of uncertainty affect many other parameters of a given system. For example, engineering design problems (Wiebenga et al. 2011), portfolio optimizations problems (Bemis et al. 2009, Fama and French 1993), etc., often have only a few primitive uncertain parameters and regression or factor models are used to obtain (3.16). For the sake of simplicity, we focus below on an individual chance constraint, so subindex k is omitted, but in §3.3.3.1 we show how the method is extended to joint chance constraints.

Eventually, our objective is to find the tightest uncertainty set $\mathcal{Z}$ such that for any feasible solution $x \in \mathbb{R}^n$ of

$$a(\zeta)^T x \le b \quad \forall \zeta \in \mathcal{Z} \tag{3.17}$$

the chance constraint

$$\Pr_{\zeta}\{\zeta : a(\zeta)^T x \le b\} \ge \beta \tag{3.18}$$

is satisfied. Constraint (3.17) is called a safe approximation of chance constraint (3.18). Furthermore, (3.17) is also the RC of the uncertain constraint with the uncertainty set $\mathcal{Z}$.

To determine $\mathcal{Z}$ and the corresponding probability bound, we first divide the domain of ζ into cells such that in each cell there are sufficient historical data. Then, using

these data, we calculate the frequency q_i of each cell $i \in V$. The true probability of a cell is denoted by p_i, and the true probability vector p is in the $(1-\alpha)$-confidence set (3.4). Let $\mathcal{Z}$ be the uncertainty set $C(S)$ given by (3.6) and $x \in \mathbb{R}^n$ be any feasible solution for the safe approximation (3.17). Then from §3.2.2 we have

$$\Pr_\zeta\{\zeta : a\,(\zeta)^T\, x \leq b\} \geq \gamma(S,\alpha) \tag{3.19}$$

with confidence level $(1-\alpha)$, where $\gamma(S,\alpha)$ is given by (3.7). The aim is to find a tight S that approximates the uncertainty region by $C(S)$ such that $\gamma(S,\alpha) \geq \beta$, and hence (3.18) holds with a $(1-\alpha)$ confidence level.

In the following section, we present an algorithm that finds such a tight uncertainty set for a given probability bound β.

3.3.2 Algorithm

In this section, we present an algorithm that iteratively constructs an uncertainty set $\mathcal{Z}$ that satisfies the probability bound β given in (3.18). We illustrate our approach using the following toy problem:

$$\max_{\boldsymbol{x}\geq 0} \ \left\{x_1 + x_2 \ : \ \Pr_\zeta\left\{\zeta \in [-1,1]^2 : \zeta_1 x_1 + \zeta_2 x_2 \leq 1\right\} \geq \beta\right\}, \tag{3.20}$$

where β is the prescribed probability, and $\zeta_1 \in [-1,1]$ and $\zeta_2 \in [-1,1]$ are the primitive uncertain parameters that we have historical data on. Later in this section, we adopt the general notation in §3.3.1, i.e., also for the toy problem, and the steps of the algorithm are explained in detail below.

Step 0. We scale ζ to $[-1,1]^\ell$, where the uncertain parameter is equivalent to the following vector:

$$a(\zeta) = a^0 + \sum_{j=1}^{\ell} \zeta_j a^j. \tag{3.21}$$

For the toy problem, ℓ is equivalent to 2, $a^0 = 0$, and $a^j \in \mathbb{R}^2$ equals the unit vector e_j; hence, $a(\zeta) = \zeta$.

Then, we calculate the frequency q_i of each cell $i \in V$ as described in §4.11. Figure 4.1 shows the historical data on ζ for the toy problem, as well as the cells that include the data. The size of the cells is such that each cell includes "enough" data, i.e., at least five observations according to a rule of thumb.

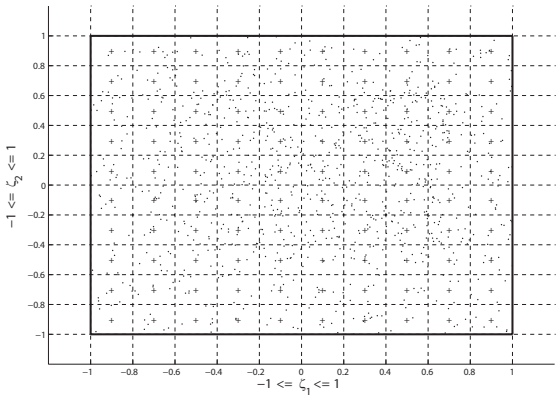


Figure 3.1 – Historical Data for ζ_1 and ζ_2

Remark 2 *Cells with low frequencies can be combined to get "enough" data. Figure 4.1 presents the standard situation where all cells have the same geometry.*

Step 1. The robust counterpart problem given by

$$\begin{aligned} &\max_{x\in\mathbb{X}} \; c^{\mathrm{T}}x \\ &\text{s.t. } a(\zeta)^{\mathrm{T}}x \leq b \quad \forall \zeta \in \mathcal{Z} \end{aligned} \tag{3.22}$$

is solved, where $\mathcal{Z}$ equals the ball-box uncertainty set:

$$BB_{\Omega} := \left\{\zeta \in \mathbb{R}^{\ell} : ||\zeta||_2 \leq \Omega, ||\zeta||_{\infty} \leq 1\right\}. \tag{3.23}$$

The exact formulation of constraint (3.22) for $\mathcal{Z} = BB_{\Omega}$, is equivalent to:

$$\begin{aligned} &z_j + w_j = -[a^j]^T x, \quad \forall j \in \{1, ..., \ell\} \\ &\sum_{j=1}^{\ell} |z_j| + \Omega\sqrt{\sum_{j=1}^{\ell} w_j^2} \leq b - [a^0]^T x, \end{aligned} \tag{3.24}$$

where z and w are additional dual variables; for the detailed derivation of the RC see Appendix 3.6. Note that the above formulation can easily be reformulated as an SOCP. In Figure 3.2, we illustrate the uncertain constraint in the toy problem, when x is fixed to the robust optimal solution x^* and $BB_{0.5}$ is the uncertainty set used in the robust counterpart.

Remark 3 *Instead of an ellipsoidal uncertainty set, we can also use other uncertainty sets such as the box. In §3.3.3.3, we discuss that in detail.*

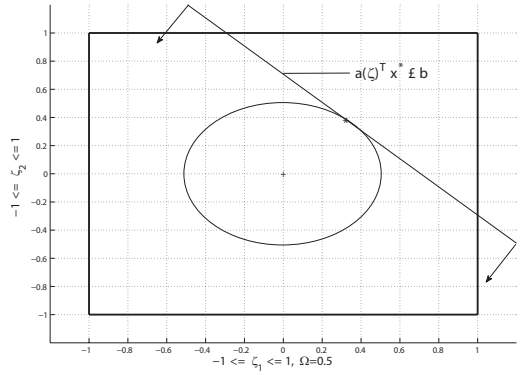


Figure 3.2 – Uncertain Constraint and Ball-Box Uncertainty Set

Step 2. We calculate the set of cells

$$S = \left\{ i \in V : a(c^i)^T x^* \leq b \right\}, \tag{3.25}$$

where $c^i = (c_1^i, c_2^i \ldots, c_\ell^i)$ is the center point of cell $i \in V$. If the center point of a cell satisfies the constraint in (3.25) for a given x^*, then we assume that all the realizations in the associated cell are feasible for the uncertain constraint. Conversely, if the center point of a cell does not satisfy (3.25) for a given x^*, then we assume that all the realizations in this cell are infeasible for the uncertain constraint. This assumption is referred as the *center point assumption* in later sections. For the toy problem, the region determined by S is presented in Figure 3.3.

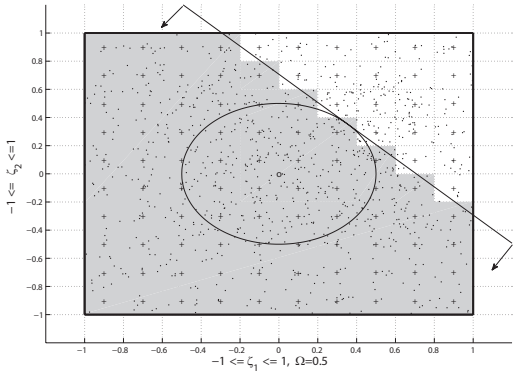


Figure 3.3 – Uncertainty Region $C(S)$

Let I be the intersection of the support, i.e., the box $[-1, 1]^\ell$, and the region determined by the constraint $[a(\zeta)^T x^* \leq b]$. Then, an important observation is that solution x^* is also robust to the uncertainty set I. In addition, the probability that

ζ is an element of I is at least the probability that ζ is an element of BB_Ω or equivalently

$$\Pr_\zeta\{\zeta \in I\} \;\geq\; \Pr_\zeta\{\zeta \in BB_\Omega\},$$

since BB_Ω is a subset of I. Hence, using I instead of BB_Ω provides a better probability bound for the optimal solution x^*. To calculate the probability bound, I is approximated by $C(S)$.

Step 3. We calculate $\gamma(S,\alpha)$ as in (3.7). If $\gamma(S,\alpha) \geq \beta$ then the region determined by I is selected as the uncertainty set and the algorithm is terminated. Otherwise, we go to Step 4.

Step 4. We increase Ω by the step size ω and go to Step 1.

Algorithm 1 (Constraint-wise algorithm)	
Inputs:	LO problem, set of cells V, frequency vector q, step size ω, confidence level $(1-\alpha)$, probability bound β, and $\Omega = 0$.
Outputs:	Uncertainty set $\mathcal{Z}$, robust optimal solution x^*, and radius Ω.
Step 0:	Scale ζ to $[-1,1]^\ell$ where the uncertain parameter is $a(\zeta) = a^0 + \sum_{j=1}^{\ell} \zeta_j a^j$.
Step 1:	Solve the robust counterpart of the given problem according to the uncertainty set BB_Ω and find the optimal solution x^*.
Step 2:	Calculate $S = \{i \in V : a(c^i)^T x^* \leq b\}$.
Step 3:	Calculate $\gamma(S,\alpha)$ **if** $\gamma(S,\alpha) \geq \beta$ **then** $\mathcal{Z} = \{\zeta \in [-1,1]^\ell : a(\zeta)^T x^* \leq b\}$ and **terminate** the algorithm **else go to** Step 4.
Step 4:	Set $\Omega = \Omega + \omega$ and **go to** Step 1.

Remark 4 *Notice that $\gamma(S,\alpha)$ is not necessarily increasing in Ω.*

We now discuss the complexity and the optimality of Algorithm 1.

Complexity. In an ℓ-dimensional uncertainty space, Ω can be at most $\sqrt{\ell}$ since BB_Ω is equivalent to the support, $[-1,1]^\ell$, when Ω is at least $\sqrt{\ell}$. Hence, Ω is changed in at most $O(\omega^{-1}\sqrt{\ell})$ iterations of the algorithm.

Optimality. 1) With probability α, the feasible set and consequently the optimal solution of the algorithm may not be contained in the feasible set of the chance constraint. The α value is selected as 0.001 in the numerical experiments. **2)** It is important to point out that, for the given inputs, another uncertainty set that is still satisfying the chance constraint may yield a better objective value for the safe approximation than that of the algorithm. One approach to find the optimal objective value for the given inputs is to start from the full space of uncertainty and to remove cells one-by-one. This means we remove the cell that yields the highest objective improvement at each iteration of the algorithm. Such a "one-by-one" removal process significantly increases the number of iterations. The numerical results in Appendix 3.8.2 show that our algorithm and the one-by-one approach yield "almost" the same objective values. **3)** In some instances, the feasible set of the safe approximation method can be empty, which means that the RC problem (3.24) is infeasible for the ball-box uncertainty set (3.23). The theory still holds since the empty set is included in the feasible set of the chance constraint but the safe approximation has no practical meaning. Note that the well-known safe approximation of Ben-Tal and Nemirovski (2000) suffers from the same phenomenon. One example for such a situation is: $[\zeta^T x \leq b \quad \forall \zeta \in BB_\Omega]$, where $b < 0$. To the best of our knowledge, for the given instance there is no way to fix the infeasibility of the safe approximation method that is proposed by Ben-Tal and Nemirovski; since $\zeta = \mathbf{0}$ is infeasible for the uncertain constraint and the ball-box uncertainty set has to be centered at $\mathbf{0}$. We propose two variants of our safe approximation method to fix the associated problem (i.e., the infeasibility of RC for BB_Ω). In the first variant, the new ball-box uncertainty is defined as $[BB_\Omega(\bar{c}) := \{\zeta \in \mathbb{R}^\ell : ||\zeta - \bar{c}||_2 \leq \Omega, ||\zeta||_\infty \leq 1\}]$, where $\bar{c}$ denotes a feasible center of the ball. We derive the RC problem according to $BB_\Omega(\bar{c})$ and apply the algorithm; for the detailed derivation of the RC see Appendix 3.6. In the second variant, we apply our algorithm for each cell separately and, in the end, we select the one that yields the uncertainty set satisfying the given probability bound β with the best objective value. The first variant requires less iteration compared to the second one but the solution is dependent on the selection of the center $\bar{c}$. On the other hand, the second variant guarantees the uncertainty set yielding the best objective value. Finally, we would like to point out that due to scaling—given by (3.16)—of the uncertain parameter, most of the time in practice we do not face the infeasibility issue that is addressed in this section. Similarly, in the practical examples at §3.4 we do not observe infeasibility of RCs thus the two variants are not required and Algorithm 1 works well.

3.3.3 Extensions

3.3.3.1 Safe Approximation of Joint Chance Constraint

Our approach can also be used to approximate a joint chance constraint:

$$\Pr_{\zeta}\left\{\zeta : a_k\left(\zeta\right)^T x \leq b_k \ \ \forall k \in \{1,\ldots,K\}\right\} \geq \beta, \tag{3.26}$$

where $x \in \mathbb{R}^n$ and k denotes the constraint index. The only difference is that we work with multiple constraints rather than a single one. We can use the same algorithm for the joint version by applying the following slight change in Step 2 of Algorithm 1:

$$\textbf{Step} \ \ \mathbf{2'}: \quad \text{Calculate } S := \{i \in V : a_k(c^i)^T x^* \leq b_k \ \ \forall k \in \{1,...,K\}\} \ . \tag{3.27}$$

Note that $C(S)$ now coincides with the region determined by all K constraints and the probability $\gamma(S,\alpha)$ calculated by the algorithm is a joint probability bound satisfied by the approximation of the given joint chance constraint. In Figure 3.4, we illustrate a $C(S)$ that is determined by multiple constraints including nonlinear ones.

Remark 5 *If we have separate chance constraints rather than a joint one, then in this case the uncertainty set of each constraint must be considered separately. Our approach can also be adapted to this case, however we do not consider that in the context of this chapter. We see the joint chance constraint as a practically and theoretically more interesting topic to look at.*

Remark 6 *Applying the Bonferroni approach to a joint chance constraint, is known to be too pessimistic (Chen et al. 2010, Chen et al. 2007, Nemirovski and Shapiro 2006).*

3.3.3.2 Extension to Nonlinear Inequalities

Our approach can be extended to nonlinear inequalities. We can focus w.l.o.g. on a single nonlinear constraint, and the robust counterpart of the uncertain constraint with the uncertainty set $\mathcal{Z}$ is given by

$$f(a(\zeta), x) \leq b \quad \forall \zeta \in \mathcal{Z}, \tag{3.28}$$

where function $f(a(\zeta), x)$ denotes the uncertain nonlinear left-hand side of the constraint. Nonlinearity may be in terms of the decision variables $x \in \mathbb{R}^n$ and/or the uncertain parameters $\zeta \in \mathbb{R}^\ell$. We have no assumption on the decision variables x as long as (3.28) is tractable; the tractable formulations of such problems are studied

in Ben-Tal et al. (2011), but we assume f is convex in the uncertain parameters ζ for any x. If this assumption holds, then only Step 2 of Algorithm 1 changes slightly as follows:

$$\textbf{Step } \mathbf{2''}: \quad \text{Calculate } S := \{i \in V : f(a(c^i), x^*) \leq b\}, \tag{3.29}$$

where x^* is the optimal solution of the robust counterpart problem with constraint (3.28). Note that the algorithm can be extended to joint nonlinear constraints with the following change:

$$\textbf{Step } \mathbf{2'''}: \quad \text{Calculate } S := \{i \in V : f_k(a_k(c^i), x^*) \leq b_k \ \ \forall k \in \{1, ..., K\}\},$$

where k denotes the constraint index. In Figure 3.4, we illustrate an iteration of the algorithm for a problem that has one linear and two nonlinear constraints in a two-dimensional uncertainty space. Note that the dark region denotes $C(S)$ and the linear constraint is presented by the dashed line.

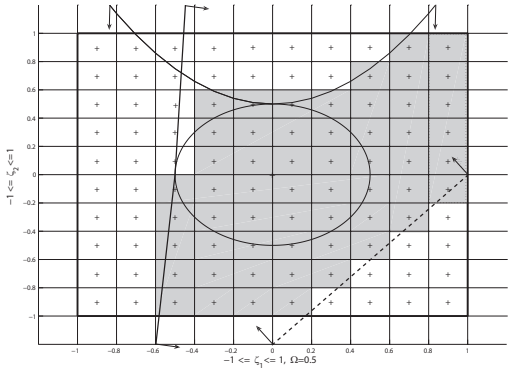


Figure 3.4 – Iteration with Nonlinear Constraints

Remark 7 *In Figure 3.4, the linear uncertain constraint is not tangent to the ellipsoidal uncertainty set $BB_{0.5}$ for the robust optimal solution. This is because the associated constraint in the RC is not binding at optimality.*

3.3.3.3 Extension to Box Uncertainty Set

So far, we have constructed a tight uncertainty set for a given uncertain optimization problem by using ellipsoids. In this subsection, we discuss how we can apply the same method starting from an uncertainty set different than the ellipsoid.
To begin with, we want the tractability of the RC to be as good as that with the ellipsoid. This is why we consider the box uncertainty set as a good choice, and in the sequel of this section we consider

$$\text{Box}_\Omega := \left\{\zeta \in \mathbb{R}^{\ell} : ||\zeta||_\infty \leq \Omega\right\}$$

as the starting uncertainty set at each iteration of our approach. The RC is no longer an SOCP given by constraints (3.24) in Algorithm 1, but it is equivalent to

$$(a^0)^T x + \Omega \sum_{j=1}^{\ell} \left|(a^j)^T x\right| \leq b, \tag{3.30}$$

that can easily be reformulated as a LO problem. Hence Step 1 of Algorithm 1 changes slightly as follows:

Step 1′ : Solve the RC with constraint (3.30) for given Ω and find the optimal x^*.

The numerical results in Appendix 3.8.3 show that using the box or the ellipsoid yields similar results in the safe approximation method. However, we have the impression that, especially for joint constraints, the ellipsoidal uncertainty set has more flexibility than the box in finding the final tight uncertainty set. This is because of the special geometry of the ellipsoid that avoids that the worst-case realizations of the uncertain parameters are in the corners of the box.

3.4 Experimental Results

In this section, we provide the results of the experiments we have conducted for the algorithm presented in §3.3.2 and its extension in §3.3.3. All computations are done on a 32-bit Windows machine equipped with a 2.66 GHz Intel Core 2 Quad processor with 3.2 GB of RAM. To solve the mathematical programs, we have used KNITRO 7.0 embedded in MATLAB 2011b. In the first experiment, we solve a simple uncertain linear optimization problem with a single constraint. The performance of our algorithm is compared with the approximation of the chance constraint presented in Ben-Tal et al. (2009). The optimal objective value is considered as the main performance measure in this experiment. Similar to the first experiment, in Appendix 3.8.1 we compare our algorithm with the randomized approach that is proposed by Calafiore and Campi (2005). In the second experiment, we focus on a robust response model of a cross-docking distribution center in China. The related robust counterpart is a nonlinear optimization problem (NLP). Furthermore, we have also used dependent data in this experiment. Finally, in the last experiment, we apply our approach to another real-life problem originated by the need of a Dutch based electronics company. The related problem, *TV tube problem*, has six uncertain parameters and many uncertain constraints. Numerical results show that our approach provides significant improvements to the nominal case of the associated problem. Last but not least, in Appendix 3.8.3 we provide an additional experiment on a variant multi-period of work scheduling problem. In this experiment we consider the extension in §3.3.3.3.

3.4.1 Illustrative Example

Similar to the example in §3.3.2, we focus on a simple linear uncertain optimization problem with an individual chance constraint. The problem is as follows:

$$\begin{aligned} \text{(M)} \quad \max \quad & x_1 + x_2 \\ \text{s.t.} \quad & \Pr_\zeta\{\zeta \in [-1,1]^2 : (1+\zeta_1)x_1 + (1+\zeta_2)x_2 \leq 10\} \geq \beta \\ & x_1, x_2 \geq 0, \end{aligned} \tag{3.31}$$

where $\zeta_1 \in [-1,1]$ and $\zeta_2 \in [-1,1]$ are the independent uncertain parameters, and β is the prescribed probability bound. In addition, we have historical data for both of the uncertain parameters separately, and each data set has a sample size of 100. To obtain the frequencies, we divide the domain of each parameter into ten equal intervals of size 0.2 such that we have enough data points in each interval. The frequencies of the parameters according to the given data are presented in Table 3.2.

Table 3.2 – Frequencies of ζ_1 and ζ_2

ζ_1, ζ_2	[-1 -0.8]	[-0.8 -0.6]	[-0.6 -0.4]	[-0.4 -0.2]	[-0.2 0]	[0 0.2]	[0.2 0.4]	[0.4 0.6]	[0.6 0.8]	[0.8 1]
freq.(ζ_1)	0.05	0.05	0.1	0.1	0.15	0.15	0.15	0.15	0.05	0.05
freq.(ζ_2)	0.025	0.075	0.2	0.15	0.05	0.125	0.175	0.1	0.075	0.025

The joint uncertainty set of ζ_1 and ζ_2 has 100 (10×10) cells and the frequency of a cell is found by multiplying the frequencies of the associated intervals for ζ_1 and ζ_2. Note that this may be done since ζ_1 and ζ_2 are independent.

The aim of the experiment is to compare the optimal objective values of our safe approximation method to those provided by the safe approximation of the chance constraint (ACC) presented in §2 of Ben-Tal et al. (2009). The individual chance constraint (3.31) is approximated by both approaches for different values of the probability bound β and numerical results are listed in Table 3.3.

In this experiment, we have used Algorithm 1 for the case of independent uncertain parameters. We use χ^2-distance as the ϕ-divergence function when $\alpha = 0.001$; see subproblem (P) in §3.2.2. The first column in Table 3.3 presents the probability bounds β and the second column gives the bound satisfied by the algorithm, where $\gamma(S, \alpha)$ represents the optimal objective value of subproblem (P), or equivalently (LD). The third column presents the radius of the minimal ball in the tight uncertainty region calculated by the algorithm. The fourth column gives the probability

Table 3.3 – Results for Example 4.1

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.	Ω_{acc}	Obj_{acc}.	%Improv.
Nom.	0.5	0	0	45	10	-	-	-
0.6	0.6	0.15	0.03	36	9.04	1.35	5.11	80.8
0.7	0.7	0.29	0.15	28	8.29	>1.41	5	65.9
0.8	0.87	0.57	0.35	15	7.12	>1.41	5	42.5
0.9	0.92	0.71	0.64	10	6.65	>1.41	5	33.1
0.91	0.92	0.71	0.64	10	6.65	>1.41	5	33.1
0.92	0.92	0.71	0.64	10	6.65	>1.41	5	33.1
0.93	0.96	0.85	0.72	6	6.24	>1.41	5	24.9
0.94	0.96	0.85	0.72	6	6.24	>1.41	5	24.9
0.95	0.96	0.85	0.72	6	6.24	>1.41	5	24.9
0.96	0.976	0.99	0.89	3	5.88	>1.41	5	17.6
0.97	0.976	0.99	0.89	3	5.88	>1.41	5	17.6
0.98	0.984	1.14	0.95	1	5.53	>1.41	5	10.7
0.99	1	1.28	1	0	5	>1.41	5	0
FRC	1	>1.41**	1	0	5	-	-	-

$* \ \alpha = 0.001$, $** \ \sqrt{2} \approx 1.41$

bound provided by the algorithm, if we would have used the ball-box as the final uncertainty set. The fifth column gives the number of cells removed from the uncertainty space to obtain S, and the sixth column presents the optimal objective value provided by our algorithm. The seventh column corresponds to the radius of the ball, which is equivalent to $\sqrt{2|\ln(1-\beta)|}$ by ACC Ben Tal et al. (2009), and the eighth column lists the associated optimal objective value. Finally, the ninth column gives the percentage of improvement in the optimal objective value of ACC when our algorithm is used, or equivalently $((\text{Obj} - \text{Obj}_{acc})/\text{Obj}_{acc}) \times 100$. ACC yields the same optimal solution when Ω is higher than $\sqrt{2}$ since the ball becomes larger than the box uncertainty set in the two-dimensional space. Hence, the uncertainty set BB_Ω in (3.23) coincides with $[-1,1]^2$ that results in the worst-case objective value of 5 for (M).

The first row in Table 3.3 is the nominal problem. We provide the tightest uncertainty set and the probability bound satisfied by the nominal solution. The last row corresponds to the worst-case solution with respect to the full space of uncertainty (FRC). The results in Table 3.3 reveal that our approach outperforms ACC with respect to the optimal objective value for the given probability bounds; also

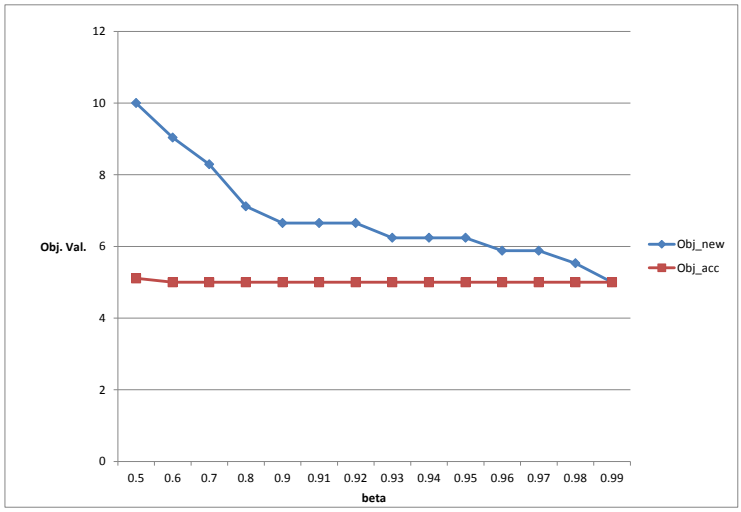


Figure 3.5 – The diamond points in the graph represent the objective values of the new approach at the associated β values (given in the horizontal axis), and the square points represent those for the classical approach. Notice that the higher the objective value is the better, since it is a maximization problem.

see Figure 3.5. For instance, when the probability bound is 0.8, the improvement in the objective value is 42.5%. Even for high probability bounds such as 0.97 our algorithm yields a 17.6% improvement in the objective. It is clear that both the improvement in the objective value and the number of cells removed from the initial uncertainty set increase as the probability bound β decreases. Furthermore, if we compare the values in the second and the fourth columns, it is easy to see that the final uncertainty set $\mathcal{Z}$ yields significantly better probability bounds then the starting ball-box uncertainty set, BB_Ω, especially when Ω is low. As a concluding remark, we have also conducted the same experiment when different ϕ-divergence functions such as Hellinger and Kullback-Leibler distances are used in subproblem (P); see the numerical results in Appendix 3.7.1.

3.4.2 Optimization of Cross-Docking Distribution Center

Our method can also be applied to the area of robust optimization via (computer) experiments. For a detailed treatment, see Myers and Montgomery (1995). The problem is to find settings for a number of design variables ($x \in \mathbb{R}^n$) such that a given objective is optimized and the performance constraints are met with a prescribed probability. One has to work with probabilities since uncontrollable noise factors ($\zeta \in \mathbb{R}^m$) influence the performance. Using (computer) experiments in which both the design variables and the noise factors are varied, response functions (or metamodels),

$\hat{y}_i(x, \zeta)$, can be developed. The constraint now becomes

$$\Pr_\zeta\{\zeta \in [-1, 1]^2 : \hat{y}_i(x, \zeta) \leq \theta_i, \ \forall i\} \geq \beta.$$

One commonly followed approach is to replace each constraint by

$$\mathbb{E}_\zeta[\hat{y}_i(x, \zeta)] + \kappa\sqrt{\mathbb{V}_\zeta[\hat{y}_i(x, \zeta)]} \leq \theta_i,$$

where κ is such that $\Pr(X \leq \kappa) \geq \beta$, where X is a standard normally distributed variable. For a recent real-life application see Wiebenga et al. (2011). A disadvantage of this approach is that one has to assume that ζ is normally distributed with a known mean and variance. Second, one has to assume that $\hat{y}_i(x, \zeta)$ is normally distributed, which is probably not the case when $\hat{y}_i(x, \zeta)$ is nonlinear in ζ.

In this example we focus on the robust response model of a cross-docking distribution center (CDDC); see Shi (2011). The associated research is motivated by the desire of a third-party logistics company to improve its supply chain management. As background information, the company distributes units from part suppliers to an assembly plant that manufactures automobiles. There are five decision factors and two environmental factors affecting the system. The environmental factors are primitive sources of the uncertainty; they are the quantity variability and the suppliers' production interruption probability. The decision factors are the number of receiving doors, shipping doors, forklifts, conveyors, and threshold parts; these factors are under the control of the users. Note that the decision factors are denoted by the coded variables $x_i \in [-1, 1]$, $i \in \{1, \ldots, 5\}$; the environmental factors are denoted by $\zeta_j \in [-1, 1]$ where $j \in \{1, 2\}$.

Because of an estimated demand growth rate of 10% to 15%, a new assembly plant will be established. When the two assembly plants operate simultaneously, the CDDC will not be able to maintain a steady distribution to the assembly plants. Therefore, the CDDC's internal operations must be optimized to satisfy the assembly plants' demand under supply uncertainty. Based on simulation results, Shi (2011) derives response functions of the performance measures to be used in the mathematical optimization problem. These measures are the dwelling time in the temporary storage area, the total throughput of the CDDC, and quantities that exceed the threshold time in the temporary storage area. We focus on the following chance constrained

problem:

$$\begin{aligned} \text{(TPL)} \quad \max \ & \mathbb{E}_\zeta \left[TT(x,\zeta) \right] \\ \text{s.t.} \ & \Pr_\zeta \{ \zeta \in [-1,1]^2 : \hat{y}_{DT}(x,\zeta) \leq 20, \ \hat{y}_{ET}(x,\zeta) \leq 40000 \} \geq \beta \end{aligned} \tag{3.32}$$

$$-1 \leq x_i \leq 1 \ \ \forall i \in \{1,\ldots,5\}, \tag{3.33}$$

where $\hat{y}_{DT}$ and $\hat{y}_{ET}$ are the response functions of dwelling time and quantities exceeding the threshold time, respectively, and $\mathcal{Z}$ is the uncertainty set. The response functions are polynomials in x_i but linear in terms of the uncertain parameters ζ_i. For complete formulas of the response functions see Appendix 3.7.2.1. The objective of (TPL) is to minimize the expected total throughput denoted by $\mathbb{E}_\zeta[TT(x,\zeta)]$, and (TPL) is a nonlinear problem since the response functions are nonlinear in x. We apply our safe approximation method to find an uncertainty set $\mathcal{Z}$ such that for any feasible solution $x \in \mathbb{R}^5$ of the RC:

$$\hat{y}_{DT}(x,\zeta) \leq 20, \ \ \forall \zeta \in Z \tag{3.34}$$

$$\hat{y}_{ET}(x,\zeta) \leq 40000, \ \ \forall \zeta \in Z, \tag{3.35}$$

the joint chance constraint (3.32) is satisfied for the given probability bound β.

Similarly to the earlier experiments, the uncertainty space is divided into 100 (10×10) cells. Furthermore, the uncertain parameters ζ_1 and ζ_2 are assumed to be independent and normally distributed in Shi (2011). These assumptions are not essential for our approach, but we have used them for the sake of comparison. Thus, random data for ζ_1 and ζ_2 are obtained from $N(20,5)$ and $N(0.02,0.01)$ with a sample size of 1000, respectively, and scaled to the interval $[-1,1]$. Table 3.4 presents the results of the experiment.

The optimal objective value of the nominal problem is 496597. Moreover, the probability bound satisfied by this solution is 0.49. In other words, the joint uncertain constraint will not be satisfied with 51% probability, when x is fixed to the nominal solution in (3.32).

The target expected total throughput of the company is 480000 (Shi 2011). Our results in Table 3.4 show that this target can be satisfied for a probability bound as high as 0.81. In addition, the immunity to 81% of the uncertainty is significantly better than that provided by the nominal solution. Between the nominal solution and the solution satisfying a bound of 0.8, the optimal objective value decreases by 3%,

Table 3.4 – Results for CDDC Example

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.49	0	0	50	496597
0.6	0.62	0.1	0^{**}	44	491096
0.7	0.72	0.18	0.21	40	486534
0.8	0.81	0.27	0.21	36	481507
0.9	0.9	0.44	0.61	31	472870
0.91	0.92	0.46	0.61	29	471925
0.92	0.92	0.46	0.61	29	471925
0.93	0.94	0.51	0.74	23	469386
0.94	0.94	0.51	0.74	23	469386
0.95	0.95	0.56	0.74	20	467129
0.96	0.96	0.62	0.83	18	464540
0.97	0.97	0.66	0.83	15	462884
0.975	0.98	0.71	0.9	12	460890
0.98	0.98	0.76	0.91	10	458977
0.99	0.99	0.86	0.94	6	455381
FRC	1	$>1.41^{***}$	1	0	445172

$*\ \alpha = 0.001$, $**\ 2 \times 10^{-8} \approx 0$, $***\ \sqrt{2} \approx 1.41$

while there is a 32% increase in the immunity to uncertainty. On the other hand, for probability bounds above 0.9, we can no longer satisfy the target. For instance, our optimal solution can not satisfy 5% of 480000, when the prescribed probability is 0.99.

The trade-off between the probability guarantee and the optimal objective value is clear in the reported results. Using the solutions in Table 3.4, the decision maker can select the best strategy for the new distribution system. This could involve accepting a small reduction from the expected target for the sake of a higher probability guarantee, or satisfying the target with a lower guarantee.

Dependent Data. Later in this example, we use the dependent data that is presented in Table 3.15; see Appendix 3.8.4. The data is obtained using a bivariate normal distribution by post-processing the "tail" cells that have less observations. The values in Table 3.15 correspond to the number of observations in the associated cells and the sample size is 3033, hence the frequency of a cell can be calculated by dividing the number of observations in the associated cell to the sample size. The total number of cells is again 100.

Table 3.5 – Results for Dependent Data

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.36	0	0	50	496597
0.6	0.62	0.24	0.06	37	483161
0.7	0.71	0.45	0.26	30	472397
0.8	0.81	0.51	0.36	23	469387
0.9	0.90	0.71	0.52	12	460890
0.91	0.91	0.75	0.53	11	459354
0.92	0.93	0.81	0.57	8	457143
0.93	0.94	0.86	0.57	6	455382
0.94	0.95	0.99	0.74	3	451126
0.95	0.95	1.04	0.8	1	449603
0.96	1	1.09	0.81	0	448139
0.97	1	1.09	0.81	0	448139
0.98	1	1.09	0.81	0	448139
0.99	1	1.09	0.81	0	448139
FRC	1	$>1.41^{**}$	1	0	445172

$*\ \alpha = 0.001$, $**\ \sqrt{2} \approx 1.41$

According to the given data, we apply our safe approximation method to the CDDC problem and the numerical results are reported in Table 3.5. The uncertainty sets that are reported in Table 3.5 are larger than the ones provided in Table 3.4. This is because of three reasons: First is the data structure, e.g., extensive data locate on the corners of the uncertainty region, namely, the top-left and the bottom-right corners in Table 3.15. Second, the ρ value in constraint (3.8) increases, since the degrees of freedom increases. Note that the degrees of freedom is 99 for the dependent case; whereas it is 81 for the independent case. Third, the sample size of the dependent data is smaller than that of the independent data. As a result, to satisfy the same probability guarantees we require larger uncertainty sets. Note that a larger uncertainty sets implies a conservative RC and this is why the optimal objective values in Table 3.5 are lower than the ones in Table 3.4. Nevertheless, we still have significant improvements to the nominal solution. For instance, the solution satisfying a bound of 0.6 has 26% higher immunity to uncertainty than that of the nominal solution and it is a considerable improvement for a 2.7% loss in the optimal objective value. For probability bounds that are higher than 0.95, the safe approximation method finds the same tight uncertainty set yielding the probability bound of one (using the discretization and the *center point assumption* of the safe approximation method).

Furthermore, the optimal objective value of the RC with $BB_{1.09}$ is 0.6% higher than the worst-case optimal 445172 (FRC).

To conclude, it is clear that using the safe approximation method yields significant improvements to the immunity to uncertainty, provided by the nominal solution, for relatively small losses in terms of the optimal objective value.

3.4.3 Optimizing Color Picture Tube

In the manufacturing process of a standard television, the *color picture tube* is assembled to the other components using a manufacturing oven. The oven temperature causes thermal stresses on different points of the tube and if the temperature is too high, it will scrap the tube due to implosions. Figure 3.6 taken from Hertog and Stehouwer (2002) gives an example of a temperature profile on a tube.

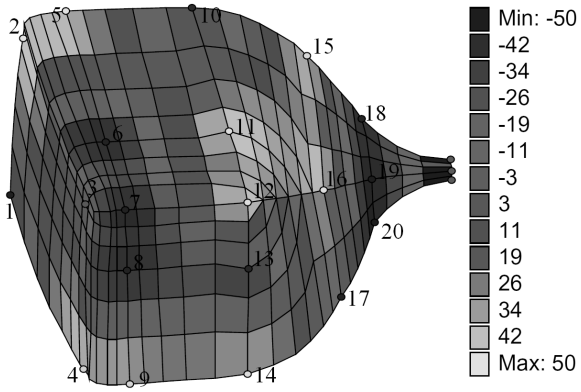


Figure 3.6 – Temperature Profile

To minimize the cost and hence the number of scraps, the manufacturer would like to make an optimal temperature profile such that the temperatures are in the specified range, the temperature differences between near locations are not too high and the maximal stress for the TV tube is minimal. Den Hertog and Stehouwer (2002) formulated the associated problem as follows:

$$
\begin{aligned}
\min\ & s_{\max} \\
\text{s.t. } & a_k + b_k^T x - s_{\max} \leq 0 && \forall k \in \{1, \ldots, K\} && (3.36)\\
& -\triangle T_{\max} \leq Ax \leq \triangle T_{\max} && && (3.37)\\
& l \leq x \leq u, && && (3.38)
\end{aligned}
$$

where $s_{\max} \in \mathbb{R}$ is the maximal stress, $a_k + b_k^T x \in \mathbb{R}$ is the stress at location k, i.e., linear in x, and $x \in \mathbb{R}^n$ represents the vector of temperatures. The vectors $l \in \mathbb{R}^n$

and $u \in \mathbb{R}^n$ are the lower and upper bounds of the decision variables, respectively. The parameter $\triangle T_{\max} \in \mathbb{R}^d$ represents the maximal allowed temperature on d location combinations. $A \in \mathbb{R}^{d \times n}$ coincides with the coefficients in the linear constraints that enforce the specified temperatures do not differ more than $\triangle T_{\max}$. There are 20 temperature points on the TV tube and hence $n = 20$; see Figure 3.6. Furthermore, these temperatures result in 216 thermal stresses on different parts of the tube so $K = 216$. The response functions of the thermal stresses, $a_k + b_k^T x$, are derived by using FEM simulator and regression in Den Hertog and Stehouwer (2002). In this example, we use the same response functions, but the decision variable x_i is replaced by $x_i(1 + \zeta_j)$, where ζ_j is the multiplicative uncertain parameter, i.e., commonly referred as the implementation error (e.g., $\zeta_j = 0.2$ means 20% implementation error in x_i).

According to the proximity of the temperature points, we form the following six subgroups:

j	1	2	3	4	5	6
$T(j)$	{1}	{2,5,10}	{3,6,7,8}	{4,9,14}	{11,12,13}	{15,16,17,18,19,20}

$T(j)$ denotes the set of indices of the decision variable(s) that are assumed to be affected by the same uncertain parameter ζ_j. This is a valid assumption since closer points in the TV tube have similar temperatures in practice. Eventually, using the safe approximation method, our objective is to find the tightest uncertainty set $\mathcal{Z}$ for the RC:

$$\begin{aligned} \min \ & s_{\max} \\ \text{s.t.} \ & a_k + \sum_{j=1}^{6} \sum_{i \in T(j)} b_{ik} x_i (1 + \zeta_j) - s_{\max} \leq 0 \quad \forall k \in \{1, \ldots, K\}, \forall \zeta \in \mathcal{Z} \\ & (3.37), (3.38), \end{aligned} \tag{3.39}$$

such that the joint chance constraint:

$$\Pr_\zeta \Big\{ \zeta \in [-1, 1]^6 : a_k + \sum_{j=1}^{6} \sum_{i \in T(j)} b_{ik} x_i (1 + \zeta_j) \leq s_{\max}, \ \forall k \in \{1, \ldots, K\} \Big\} \geq \beta \tag{3.40}$$

is satisfied for any feasible RC solution $(x, s_{\max})$, where β is the given probability bound. The RC problem has 21 decision variables including $s_{\max}$, six primitive uncertain parameters, 216 linear uncertain constraints (i.e., given by constraint (3.39)) and 56 linear constraints (i.e., given by constraints (3.37) and (3.38)).

Data. The data for implementation errors are invented by us and the data range is divided to five equal intervals. The frequencies of the associated intervals are shown in Table 3.16 and 3.17; see Appendix 3.8.4, we have the same frequencies in two different data ranges that are: [-.1, .1] and [-.2, .2]. These ranges correspond to 10% and 20% implementation errors, respectively. In addition, we assume the uncertain parameters are independent and hence the frequency of a cell may be found by multiplying the frequencies of the associated intervals for $(\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_6)$. The total number of cells in the joint uncertainty space is 15625 (5^6).

The numerical results are shown in Table 3.6 and Table 3.7. In both tables, it is easy

Table 3.6 – TV Tube Example (10% Imp. Err.)

β	$\gamma(S, \alpha^*)$	Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom	0^{**}	0	15625	14.14
0.3	0.3	0.56	12642	14.4
0.35	0.35	0.6	11967	14.42
0.4	0.41	0.64	11245	14.44
0.45	0.46	0.68	10447	14.45
0.5	0.53	0.72	9547	14.47
0.6	0.63	0.78	8031	14.5
0.7	0.71	0.84	6753	14.52
0.75	0.76	0.88	5938	14.54
0.8	0.82	0.94	4772	14.56
0.85	0.85	0.98	4046	14.58
0.9	0.91	1.06	2833	14.61
0.92	0.93	1.1	2321	14.63
0.95	0.95	1.16	1648	14.65
0.96	0.97	1.2	1300	14.67
0.97	0.97	1.22	1140	14.68
0.98	0.98	1.28	736	14.7
0.99	0.99	1.34	455	14.73
FRC	1	$>2.45^{***}$	0	14.91

$* \ \alpha = 0.001$, $** \ 2{\times}10^{-8} \approx 0$, $*** \ \sqrt{6} \approx 2.45$

to see that the nominal solution is not immune to the implementation errors. To be more precise, if the decision variables $(x, s_{\max})$ are fixed to the nominal solution in the joint chance constraint (3.40), then the left-hand side probability is almost zero (i.e., 2×10^{-8}). This means that the ζ values that are feasible for the joint

constraint, are realized with almost zero probability. This is why implementing the nominal solution can be a risky decision in practice, but using the safe approximation method we can find significantly better solutions.

The numerical results in Table 3.6 show: Between the nominal solution and the solution satisfying a bound of 0.3, the optimal objective value increases by 1.8%, while there is a 30% increase in the immunity to uncertainty. In addition, the solution satisfying a bound of 0.85 has an optimal objective value that is 3% higher than that of the nominal solution. These are significant improvements in the immunity to uncertainty for the losses in the optimal objective value.

Table 3.7 – TV Tube Example (20% Imp. Err.)

β	$\gamma(S,\alpha^*)$	Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0^{**}	0	15625	14.14
0.3	0.31	0.56	12626	14.64
0.35	0.37	0.6	11857	14.67
0.4	0.43	0.64	11015	14.7
0.45	0.46	0.66	10593	14.72
0.5	0.53	0.7	9680	14.75
0.6	0.63	0.76	8078	14.8
0.7	0.71	0.82	6640	14.85
0.75	0.75	0.86	5906	14.87
0.8	0.82	0.92	4594	14.92
0.85	0.86	0.96	3776	14.95
0.9	0.9	1.02	2827	14.99
0.92	0.92	1.06	2286	15.02
0.95	0.95	1.14	1463	15.07
0.96	0.97	1.18	1114	15.1
0.97	0.97	1.2	950	15.11
0.98	0.98	1.24	676	15.13
0.99	0.99	1.3	372	15.17
FRC	1	$> 2.45^{***}$	0	15.44

$* \ \alpha = 0.001$, $** \ 2{\times}10^{-8} \approx 0$, $*** \ \sqrt{6} \approx 2.45$

Note that when we increase the implementation errors from %10 to %20, then the variance from the nominal case increases and we require larger tight uncertainty sets to satisfy the same probability bounds. This is why the number of cells removed from the full space of uncertainty is fewer when the implementation errors are higher; see

the fourth columns of Table 3.6 and 3.7. A larger tight uncertainty set implies a more restrictive RC and hence the optimal objective values shown in Table 3.7 are on average 2.2% higher than those provided in Table 3.6. The lowest difference between two optimal objectives, i.e., 1.7%, is obtained at 0.3 probability bound, and the highest, i.e., 2.9%, is obtained at 0.99, and there is a gradual increase in between.

3.5 Conclusions

In this chapter, we have proposed new safe approximations for joint chance constraints. Using historical data and goodness-of-fit statistics based on ϕ-divergence, we constructed the uncertainty sets that are used in safe approximations. The numerical results show that our approach yields tighter uncertainty sets, and therefore better objective values than the existing method, for the same probability guarantees, especially when the number of uncertain parameters is low. In addition, we do not impose the assumptions that the uncertain parameters are independent or certain moments are known. Last but not least, the new approach can also handle nonlinear inequalities.

It is important to observe that the computational performance of our approach is highly dependent on the number of uncertain parameters. Furthermore, we may require many data points, especially when the uncertain parameters are dependent and the number of uncertain parameters is high, and this data requirement may be hard to manage in practice. In future research, we will investigate the improvement of our approach in such situations. The extension of our approach to simulation based optimization and nonlinear problems will also be further analyzed in future research.

APPENDICES

3.6 Derivation of Robust Counterpart

Let $\mathcal{Z} = \{\zeta : P_s\zeta + p_s \in K^s, s = (1, \ldots, S)\}$ where K^s denotes a closed convex pointed cone with a non-empty interior. The semi-infinite representation:

$$(a^0 + \sum_{i=1}^{\ell} \zeta_i a^i)^T x \leq b, \ \ \forall \zeta \in \mathcal{Z}$$

is equivalent to the following RC:

$$\max_{\zeta \in \mathcal{Z}} \Big\{(a^0 + \sum_{i=1}^{\ell} \zeta_i a^i)^T x\Big\} \leq b. \tag{3.41}$$

Using the Conic Duality theorem, (3.41) is equivalent to

$$\begin{aligned} &\min_{y^s \in K_*^s : s \in \{1,\ldots,S\}} \Big\{ \sum_{s=1}^{S} p_s^T y^s + (a^0)^T x\Big\} \leq b \\ &\sum_{s=1}^{S} (P_s^T y^s)_j = -(a^\ell)^T x \ \ \forall j \in \{1, \ldots, \ell\}, \end{aligned} \tag{3.42}$$

where K_*^s denotes the dual cone of K^s. For the detailed derivation of (3.42), see Theorem 1.3.4 and Corollary 1.3.6 of Ben-Tal et al. (2009). Then the minimization in (3.42) is removed—as it is often done in RO—and we have the following equivalent system of conic inequalities:

$$\begin{aligned} &\sum_{s=1}^{S} p_s^T y^s + (a^0)^T x \leq b \\ &\sum_{s=1}^{S} (P_s^T y^s)_j = -(a^\ell)^T x \ \ \forall j \in \{1, \ldots, \ell\} \\ &y^s \in K_*^s \quad \forall s \in \{1, \ldots, S\}. \end{aligned} \tag{3.43}$$

The conic representation of the uncertainty set $BB_\Omega(\bar{c})$ is equivalent to

$$\mathcal{Z} = \{\zeta : P_1\zeta + p_1 \in K^1, P_2\zeta + p_2 \in K^2\}$$

where

- $P_1\zeta \equiv [\zeta; 0], p_1 \equiv [0_{1\times\ell}; 1]$ and $K^1 := \{(e, t) \in \mathbb{R}^\ell] \times \mathbb{R} : ||e||_\infty \leq t\}$
- $P_2\zeta \equiv [\zeta; 0], p_2 \equiv [-\bar{c}; \Omega]$ and $K^2 := \{(e, t) \in \mathbb{R}^\ell \times \mathbb{R} : ||e||_2 \leq t\}$.

Then using (3.43), the RC for $BB_\Omega(\bar{c})$ is equivalent to

$$\begin{aligned} &\tau_1 - w^T\bar{c} + \Omega\tau_2 + (a^0)^T x \leq b \\ &(z+w)_\ell = -(a^j)^T x, \;\; j = (1,\ldots,\ell) \\ &||z||_1 \leq \tau_1 \;\; (K_*^1 := \{(z,\tau_1) \in \mathbb{R}^\ell \times \mathbb{R} : ||z||_1 \leq \tau_1\}) \\ &||w||_2 \leq \tau_2 \;\; (K_*^2 := \{(w,\tau_2) \in \mathbb{R}^\ell \times \mathbb{R} : ||w||_2 \leq \tau_2\}), \end{aligned} \tag{3.44}$$

where $y^1 = [z;\tau_1] \in K_*^1$ and $y^2 = [w;\tau_2] \in K_*^2$. Consequently, (3.44) is equivalent to

$$\begin{aligned} &\sum_{j=1}^{\ell} |z_j| + \Omega\sqrt{\sum_{j=1}^{\ell} w_j^2} - w^T\bar{c} \leq b - [a^0]^T x, \\ &z_j + w_j = -(a^j)^T x, \quad \forall j \in \{1,...,\ell\}\,. \end{aligned} \tag{3.45}$$

□

3.7 Data and Additional Results

3.7.1 Extra Results for Example 4.1

Table 3.8 – Kullback-Leibler Distance

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.	Ω_{acc}	$\text{Obj}_{acc.}$	%Improv.
Nom.	0.5	0	0	45	10	-	-	-
0.60	0.69	0.29	0.03	28	8.30	1.35	5.11	62
0.70	0.78	0.43	0.22	21	7.67	>1.41	5	53.4
0.80	0.86	0.57	0.35	15	7.13	>1.41	5	42.5
0.90	0.92	0.71	0.64	10	6.66	>1.41	5	33.2
0.91	0.92	0.71	0.64	10	6.66	>1.41	5	33.2
0.92	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.93	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.94	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.95	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.96	0.98	0.99	0.89	3	5.88	>1.41	5	17.6
0.97	0.98	0.99	0.89	3	5.88	>1.41	5	17.6
0.98	0.98	0.99	0.89	3	5.88	>1.41	5	17.6
0.98	0.99	1.14	0.95	1	5.54	>1.41	5	10.7
0.99	1	1.28	0.98	0	5	>1.41	5	0
1	-	-	-	5	-	-	-	

$*$ α=0.001

Table 3.9 – Hellinger Distance

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert-\lvert S\rvert$	Obj.	Ω_{acc}	Obj_{acc}.	%Improv.
Nom.	0.5	0	0	45	10	-	-	-
0.60	0.66	0.29	0.03	28	8.30	1.35	5.11	62
0.70	0.75	0.43	0.22	21	7.67	>1.41	5	53.4
0.80	0.83	0.57	0.35	15	7.13	>1.41	5	42.5
0.90	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.91	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.92	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.93	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.94	0.96	0.99	0.89	3	5.88	>1.41	5	17.6
0.95	0.96	0.99	0.89	3	5.88	>1.41	5	17.6
0.96	0.96	0.99	0.89	3	5.88	>1.41	5	17.6
0.97	0.98	1.14	0.95	1	5.54	>1.41	5	10.7
0.98	0.98	1.14	0.95	1	5.54	>1.41	5	10.7
0.99	1	1.27	0.98	0	5	>1.41	5	0
1	-	-	-	-	5	-	-	-

$*$ α=0.001

3.7.2 Example 4.3

3.7.2.1 Response Functions

$$\begin{aligned}\mathbb{E}_\zeta[TT(x,\zeta)] = &-479700 - 39819.17*x(1) - 20253.25*x(2) + 312.12*x(3) - 7339.86*x(4) - 339.78*x(5) - \\ &7895.49*x(1)*x(2) - 121.06*x(1)*x(3) - 33.75*x(1)*x(4) + 21.24*x(1)*x(5) - \\ &7.36*x(2)*x(3) + 649.55*x(2)*x(4) + 1136.31*x(2)*x(5) + 788.4*x(3)*x(4) + \\ &407.64*x(3)*x(5) - 1101.55*x(4)*x(5) + 34063.49*x(1)^2 + 17810.89*x(2)^2 + \\ &108.13*x(3)^2 + 10333.23*x(4)^2 - 1107.72*x(5)^2.\end{aligned}$$

$$\begin{aligned}\hat{y}_{DT}(x,\zeta) = &-8.57 + 1.2*x(1) + 2.04*x(2) - 0.17*x(3) + 0.78*x(4) + 3.30*x(5) - 0.44*x(1)*x(2) + \\ &0.29*x(1)*x(3) - 0.26*x(1)*x(4) + 0.33*x(1)*x(5) + 0.21*x(2)*x(3) - 0.45*x(2)*x(4) + \\ &0.55*x(2)*x(5) - 0.061*x(3)*x(4) + 0.062*x(3)*x(5) + 0.35*x(4)*x(5) - 0.63*x(1)^2 - \\ &1.27*x(2)^2 + 0.19*x(3)^2 - 0.25*x(4)^2 - 0.11*x(5)^2 + \\ &\{7.11 + 0.78*x(1) + 1.63*x(2) - 0.081*x(3) + 0.57*x(4) + 2.72*x(5)\}*\zeta_1 + \\ &\{3.21 + 0.46*x(1) + 0.49*x(2) - 0.073*x(3) + 0.16*x(4) + 1.17*x(5)\}*\zeta_2 \leq 0.\end{aligned}$$

$$\begin{aligned}\hat{y}_{ET}(x,\zeta) = &-7517.8 + 10256.36*x(1) + 13753.61*x(2) - 300.42*x(3) + 4379.24*x(4) + 52.43*x(5) + \\ &5415.96*x(1)*x(2) + 437.38*x(1)*x(3) + 214.75*x(1)*x(4) + 597.11*x(1)*x(5) - \\ &97.79*x(2)*x(3) - 1618.36*x(2)*x(4) - 724.67*x(2)*x(5) - 1639.28*x(3)*x(4) - \\ &1243.25*x(3)*x(5) + 1728.59*x(4)*x(5) - \\ &1118.43*x(1)^2 - 1072.35*x(2)^2 + 226.71*x(3)^2 - 372.2*x(4)^2 + 148.92*x(5)^2 + \\ &\{36087.44 + 13066.74*x(1) + 17605.17*x(2) - 739.11*x(3) + 5944.33*x(4) + 446.33*x(5)\}*\zeta_1 + \\ &\{-10868 - 3824.22*x(1) - 5975.83*x(2) + 209.48*x(3) - 2506.4*x(4) - 579.61*x(5)\}*\zeta_2 \leq 0.\end{aligned}$$

3.8 Additional Experiments

3.8.1 Comparison with Randomized Approach

In this experiment, we apply the randomized approach that is proposed by Calafiore and Campi (2005) to the illustrative experiment shown in §3.4.1. The objective is to compare these results to that of the safe approximation method reported in Table 3.3.

Calafiore and Campi (2005) propose a randomized algorithm for uncertain convex optimization problems. The algorithm is as follows: For an uncertain constraint $[f(x,\zeta) \leq 0]$ at hand, they randomly extract $\mathcal{N}$ realizations of the uncertain parameter (ζ) in an independent fashion using the known probability distribution $(\mathbb{P})$ of ζ. Note that each extraction $(\hat{\zeta})$ coincides with a discrete constraint $[f(x,\hat{\zeta}) \leq 0]$. Then, instead of the exact uncertain problem, they solve the one with randomly extracted $\mathcal{N}$ constraints. Eventually, the optimal solution of the randomized approach satisfies the following probability bound that is proved by Campi and Garatti (2008):

$$\Pr\{V(x^*_{\mathcal{N}}) > \epsilon\} \leq \underbrace{\sum_{i=0}^{d-1} \binom{\mathcal{N}}{i} \epsilon^i (1-\epsilon)^{\mathcal{N}-i}}_{B^{\epsilon}_{\mathcal{N}d}}, \tag{3.46}$$

where $B^{\epsilon}_{\mathcal{N}d}$ corresponds to the associated probability bound, $V(x)$ denotes the violation probability for a given solution $x \in \mathbb{R}^n$, ϵ is the given violation probability bound that is "generally" close to zero and $x^*_{\mathcal{N}} \in \mathbb{R}^n$ coincides with the optimal solution of the convex optimization problem with randomly selected $\mathcal{N}$ constraints. For a detailed overview of the randomized approach, we refer to Calafiore and Campi (2005), Calafiore (2006) and Campi and Garatti (2008). If we relate (3.46) to our notation: $\left[\Pr\left\{\Pr_{\zeta\sim\mathbb{P}}\{\zeta : f(x,\zeta) \leq 0\} \geq \beta\right\} \geq 1-\alpha\right]$, then the probability bound β of the chance constraint coincides with $1-\epsilon$ and the confidence level $1-\alpha$ corresponds to $1 - B^{\epsilon}_{\mathcal{N}d}$ (i.e., $B^{\epsilon}_{\mathcal{N}d} = \alpha$). This means we should find the closest $\mathcal{N}$ value that satisfies $B^{\epsilon}_{\mathcal{N}d} \cong \alpha$ (see Table 3.10 for the values of $\mathcal{N}$ when $d = 2$).

Notice that the probability distribution $\mathbb{P}$ is unknown in the safe approximation method proposed in this chapter. However, the disadvantage of the randomized approach is that $\mathbb{P}$ has to be known since constraints are extracted according to a known $\mathbb{P}$. Therefore, we assume that the empirical estimates, i.e., q in our notation, denote the required probability distribution $\mathbb{P}$ for the randomized approach. As a remark we have 100 cells in the uncertainty space so that each independent draw is done from 100 constraints. Notice that $\mathcal{N}$ can be larger than 100 for given β and α, and

Table 3.10 – Randomized versus Safe Approximation

($\boldsymbol{\alpha} = 0.001$)		Obj. Val.	
β	$\mathcal{N}^*$	RA**	SA
0.6	19	6.15	9.04
0.7	27	6.12	8.29
0.8	42	5.8	7.12
0.9	88	5.56	6.65
0.91	98	5.47	6.65
0.92	111	5.45	6.65
0.93	127	5.48	6.24
0.94	149	5.42	6.24
0.95	180	5.34	6.24
0.96	226	5.32	5.88
0.97	303	5.21	5.88
0.98	457	5.16	5.53
0.99	919	5	5

$*$ $d = 2$, $**$ Avg. of 20 rep.

due to randomization a constraint can be selected more than one time. Finally, for different probability bounds we compare the optimal objective value of the randomized approach to that of the safe approximation method proposed in this chapter. The numerical results are shown in Table 3.10, where the last column (SA), taken from the results of Experiment 3.4.1, is the optimal objective value of the safe approximation method. Note that for the randomized approach we report the average objective value of 20 replications in the third column of Table 3.10, i.e., denoted by RA. This is done to have a fair comparison between the two approaches.

The numerical results in Table 3.10 show that the safe approximation method performs better than the randomized approach. This is a logical outcome, since the safe approximation method is more objective oriented and takes into account the optimal solution, to construct the uncertainty set, at each iteration of the algorithm. On the other hand, the randomized approach does not consider the objective value while extracting constraints. Secondly, the numerical results also show that the difference between the two approaches decreases as the probability bound β increases. This is another expected result, since for probability bounds close to 1 the uncertainty set has to be close to the full space of the uncertainty, which means that there is no difference between the two approaches in this case.

3.8.2 Comparison with One-by-One Approach

The objective of this section is to compare the optimal objective values of the one-by-one approach (that is addressed in §3.3.2) to that of the safe approximation method proposed in the chapter. The numerical example is selected as the illustrative experiment in §3.4.1 (i.e., the results of the safe approximation method for the given probability bounds are taken from Table 3.3). The numerical results are presented in the following table.

Table 3.11 – One-by-One versus Safe Approximation

β	Obj. Val.	
	1-by-1	SA
0.92	6.65	6.65
0.93	6.25	6.24
0.94	6.25	6.24
0.95	6.25	6.24
0.96	5.88	5.88
0.97	5.88	5.88
0.98	5.55	5.53
0.99	5	5
$\alpha = 0.001$		

Note that the one-by-one approach provides the "exact" optimal solution for a given probability bound with a confidence level of $1-\alpha$, but it requires many more iterations compared to our method as the probability bound gets lower. The numerical results reveal that the safe approximation method yields objective values that are very close (or the same) to the one-by-one approach.

3.8.3 Multi-Period Work Scheduling Problem

In this experiment, we solve a modified version of multi-period work scheduling (MWS) problem. MWS is a linear optimization problem used to schedule employees for a multi-period time horizon where the demand changes over time.

Computer Service Store. CLS is a computer service store that requires the following skilled-repair times in the next five months: 3000, 3500, 4000, 4500, and 5500. The repair work is done by skilled technicians and these technicians can each work up to 160 hours per month. Furthermore, the technicians may train apprentices to meet future demand. It takes an average of 50 hours to train an apprentice, and

new technicians start serving CLS in the month following their training session. In addition, the training sessions have 100% efficiency, so an apprentice always becomes a technician at the end of the training period. The hiring of technicians is done only in the first period and the start-up cost of hiring a technician is \$8000. In addition, each technician is paid \$2000 and each apprentice costs \$1000 per month. On the other hand, 5% of the technicians quit at the end of each month. Finally, the objective of CLS is to minimize the total labor cost incurred to meet the demand in the next five months. The mathematical model of this problem is presented below:

$$
\begin{aligned}
\text{(NMWS)} \quad \min \ & \sum_{i=1}^{5} 1000x_i + \sum_{i=1}^{5} 2000y_i + 8000y_1 & & \\
\text{s.t.} \ & 160y_i - 50x_i \geq d_i & i \in \{1,\ldots,5\} \quad & (3.47)\\
& 0.95y_i + x_i = y_{i+1} & i \in \{1,\ldots,4\} \quad & (3.48)\\
& x_i, y_i \geq 0 & i \in \{1,\ldots,5\}, \quad & (3.49)
\end{aligned}
$$

where y_i represents the number of technicians, x_i corresponds to the number of apprentices in training, and d_i is the repair time demanded in period $i \in \{1,\ldots,5\}$. In practice, the average working and training hours usually deviate from the estimated values because of overtime, illness, vacations, and other factors. We have historical data for 120 months giving the average working and training hours spent per technician each month. These data are used in Table 3.12 to derive the frequencies. Note

Table 3.12 – Frequencies for Working (W.H.) and Training (T.H.) Hours

ζ_1, ζ_2	[-1 -0.8]	[-0.8 -0.6]	[-0.6 -0.4]	[-0.4 -0.2]	[-0.2 0]	[0 0.2]	[0.2 0.4]	[0.4 0.6]	[0.6 0.8]	[0.8 1]
W.H.	[120 128]	[128 136]	[136 144]	[144 152]	[152 160]	[160 168]	[168 176]	[176 184]	[184 192]	[192 200]
freq.(ζ_1)	0.02	0.04	0.1	0.1	0.2	0.3	0.1	0.1	0.02	0.02
T.H.	[30 34]	[34 38]	[38 42]	[42 46]	[46 50]	[50 54]	[54 58]	[58 62]	[62 66]	[66 70]
freq.(ζ_2)	0.015	0.07	0.1	0.15	0.15	0.17	0.15	0.11	0.07	0.015

that the working hours range from 120 to 200, so the mean is 160 and the half-length of the data range is 40. Similarly, for the training hours the mean is 50 and the half-length of the data range is 20. Using this information from the historical data, we introduce uncertainty to constraint (3.47) as follows:

$$
\begin{aligned}
\min \ & \sum_{i=1}^{5} 1000x_i + \sum_{i=1}^{5} 2000y_i + 8000y_1 & \\
\text{s.t.} \ & \Pr_{\zeta}\{\zeta \in [-1,1]^2 : (160 + 40\zeta_1)\, y_i - (50 + 20\zeta_2)\, x_i \geq d_i, \forall i \in \{1,\ldots,5\}\} \geq \beta & (3.50)\\
& (3.48), (3.49),
\end{aligned}
$$

where $\zeta_1 \in [-1,1]$ and $\zeta_2 \in [-1,1]$ are the uncertain parameters, and β is the prescribed probability bound. The frequencies of the working and training hours are

scaled into the frequencies of ζ_1 and ζ_2 in Table 3.12. Furthermore, the uncertain parameters are independent; therefore, the joint frequencies can be derived similarly to the first experiment. The joint uncertainty region is again divided into 100 (10×10) cells. Eventually, using our safe approximation method, we find the tightest uncertainty set $\mathcal{Z}$ such that for any feasible solution (x, y) of the RC:

$$(160 + 40\zeta_1)\, y_i - (50 + 20\zeta_2)\, x_i \geq d_i, \forall i \in \{1, \ldots, 5\}\,, \forall \zeta \in \mathcal{Z}, \tag{3.51}$$

the joint chance constraint (3.50) is satisfied for the given probability bound β. In this experiment we use the extension in §3.3.3.1, i.e., the safe approximation of the joint chance constraint. The results are reported in Table 3.13.

Table 3.13 – Results for CLS Example

β	$\gamma(S, \alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.49	0	0	50	448105
0.6	0.63	0.12	0^{**}	44	462691
0.65	0.66	0.14	0^{**}	43	465214
0.7	0.75	0.3	0.12	35	486434
0.75	0.77	0.32	0.12	34	489222
0.8	0.8	0.38	0.33	31	497782
0.85	0.86	0.5	0.38	25	515830
0.9	0.9	0.56	0.55	22	525351
0.91	0.91	0.58	0.55	21	528603
0.92	0.94	0.7	0.66	15	548988
0.93	0.94	0.7	0.66	15	548987
0.94	0.94	0.72	0.8	14	552538
0.95	0.95	0.74	0.8	13	556134
0.96	0.96	0.78	0.85	11	563468
0.97	0.98	0.9	0.9	5	586672
0.98	0.98	0.94	0.93	3	594834
0.99	0.99	0.98	0.95	1	603225
FRC	1	$>1.41^{***}$	1	0	621356

$*\ \alpha = 0.001$, $**\ 1.9\times10^{-8} \approx 0$, $***\ \sqrt{2} \approx 1.41$

The meanings of the columns in Table 3.13 are the same as for the first experiment. Note that the optimal objective values for the nominal problem (NMWS) and the robust counterpart for the full space of uncertainty (FRC) are 448105 and 621356, respectively (see the first and last row of Table 3.13). The results show that when

the probability bound is as low as 0.6, the optimal objective value calculated by the algorithm is 3% higher than that of the nominal solution. Moreover, with respect to the nominal solution, we see a 14% increase in the immunity to uncertainty in constraint (3.50), which is a considerable improvement for a 3% sacrifice in terms of the objective value. For the higher probability bounds of 0.92 and 0.94, the improvement in the optimal objective value of FRC is 12% and 11%, respectively. Furthermore, for the probability bound of 0.98, the algorithm improves the objective value of FRC by 4% and the solution is robust to at least 98.3% of the uncertainty. It is clear that when the probability bound β increases, we remove fewer cells from the initial uncertainty region and the radius Ω of BB_{Ω} gets larger. Ultimately, the decision maker must make the decision by looking at the results in Table 3.13 and choosing the best option for CLS.

Later in this example, we consider the extension in §3.3.3.3, i.e., using the box instead of the ellipsoid as the starting uncertainty set of our algorithm. Note that we apply this extension to the same problem. In addition, we test the new approach for the same data set and when the inputs of the algorithm such as the number of cells and step size ω are held constant. The numerical results are presented in Table 3.14. The symbol $(*)$ denotes an instance where using the box yields a better optimal objective value than using the ellipsoid for a given probability bound β.

Table 3.14 – Results for CLS Example (Box)

β	$\gamma(S, \alpha^*)$	Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.49	0	50	448105
0.6	0.63	0.1	44	461039(∗)
0.65	0.66	0.12	43	463715(∗)
0.7	0.78	0.3	33	489236
0.75	0.78	0.3	33	489236
0.8	0.80	0.34	31	495286(∗)
0.85	0.90	0.5	22	521022
0.9	0.90	0.5	22	521022
0.91	0.91	0.52	21	524423(∗)
0.92	0.96	0.7	11	557107
0.93	0.96	0.7	11	557107
0.94	0.96	0.7	11	557107
0.95	0.96	0.7	11	557107
0.96	0.96	0.7	11	557107(∗)
0.97	1	0.9	0	598402
0.98	1	0.9	0	598402
0.99	1	0.9	0	598402(∗)
FRC	1	$>1.41^{**}$	0	621356

$*$ $\alpha = 0.001$, $**$ $\sqrt{2} \approx 1.41$

The numerical results reveal that using the box or the ellipsoid as the starting uncertainty set yields similar optimal objective values, e.g, the highest difference between optimal objective values of two approaches is around 1%. Nevertheless, using the ellipsoid is more flexible in finding the final tight uncertainty sets for the CLS problem. For instance, if the probability bound is in between 0.92 and 0.96 or higher than 0.96, then the safe approximation method using the box finds only one uncertainty set for each of the cases; whereas, the results in Table 3.13 show that the safe approximation method using the ellipsoid finds a unique tight uncertainty set for each of the probability bounds (except 0.92 and 0.93).

3.8.4 Dependent Data Set and Data Set of Example 4.4

Table 3.15 – Dependent Data used in Example 4.3

		ζ_2									
	Cells	1	2	3	4	5	6	7	8	9	10
ζ_1	1	66	30	27	23	9	5	5	5	5	5
	2	53	35	38	23	8	13	5	5	5	5
	3	42	28	53	35	44	28	17	6	5	5
	4	32	35	44	79	81	49	26	28	9	5
	5	28	23	53	85	86	70	69	31	14	13
	6	15	15	41	46	83	102	83	67	23	20
	7	5	6	12	36	46	75	73	51	38	30
	8	5	8	5	14	30	49	65	43	38	42
	9	5	5	5	7	15	20	25	34	28	56
	10	5	5	5	5	5	12	15	28	15	79

(∗) $N = 3033$

Table 3.16 – Data Set 1 (10% Imp. Err.)

	[-.1 -.06]	[-.06 -.02]	[-.02 .02]	[.02 .06]	[.06 .1]
ζ_1	0.1	0.21	0.29	0.22	0.18
ζ_2	0.09	0.18	0.38	0.23	0.12
ζ_3	0.13	0.23	0.3	0.17	0.17
ζ_4	0.11	0.22	0.31	0.24	0.12
ζ_5	0.09	0.2	0.28	0.23	0.2
ζ_6	0.17	0.22	0.23	0.2	0.18

(∗) $N_j = 100\ \forall j \in \{1, \ldots, 6\}$

Table 3.17 – Data Set 2 (20% Imp. Err.)

	[-.2 -.12]	[-.12 -.04]	[-.04 .04]	[.04 .12]	[.12 .2]
ζ_1	0.1	0.21	0.29	0.22	0.18
ζ_2	0.09	0.18	0.38	0.23	0.12
ζ_3	0.13	0.23	0.3	0.17	0.17
ζ_4	0.11	0.22	0.31	0.24	0.12
ζ_5	0.09	0.2	0.28	0.23	0.2
ζ_6	0.17	0.22	0.23	0.2	0.18

(∗) $N_j = 100\ \forall j \in \{1, \ldots, 6\}$

Bibliography

Bemis, C., X. Hu, W. Lin, S. Moazeni, L. Wang, T. Wang, J. Zhang. 2009. Robust portfolio optimization using a simple factor model. IMA Preprint Series No. 2284, University of Minnesota, Minneapolis, MN.

Birge, J. R., J. R. B. Wets. 1986. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Math. Programming Stud.* **27** 54–102.

Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* **88**(3) 411–424.

Ben-Tal, A., A. Nemirovski. 2002. Robust optimization: methodology and applications. *Math. Programming* **92**(3) 453–480.

Ben-Tal, A., A. Nemirovski. 2009. On safe tractable approximations of chance-constrained linear matrix inequalities. *Math. Oper. Res.* **34**(1) 1–25.

Ben-Tal, A., D. den Hertog, J.-P. Vial. 2012. Deriving robust counterparts of nonlinear uncertain inequalities. CentER Discussion Paper No. 2012-053, Tilburg University, Tilburg, The Netherlands.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization*. Princeton Press, Princeton, NJ.

Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. 2012. Robust optimization with uncertainty regions based on ϕ-divergence. CentER Discussion Paper No. 2011-061, Tilburg University, Tilburg, The Netherlands. *Management Sci.* Forthcoming.

Burkauskas, A. 1986. On the convexity problem of probabilistic constraint stochastic programming problems. *Alkal. Matema. Lapok* **12** 77–90.

Calafiore, G. C. 2007. Ambiguous risk measures and optimal robust portfolios. *SIAM J. Optim.* **18**(3) 853–877.

Calafiore, G. C., M. C. Campi. 2005. Uncertain convex programs: randomized solutions and confidence levels. *Math. Programming* **102** 25–46.

Campi, M., S. Garatti. 2008. The exact feasibility of randomized solutions of robust convex programs. *SIAM J. Optim.* **19**(3) 1211–1230.

Charnes, A., W. W. Cooper. 1959. Chance-constrained programming. *Management*

Sci. **6**(1) 73–79.

Charnes, A., W. W. Cooper, G. H. Symonds. 1958. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Sci.* **4**(3) 235–263.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Oper. Res.* **55**(6) 1058–1107.

Chen, W., M. Sim, J. Sun, C. P. Teo. 2010. From CVaR to uncertainty set: implications in joint chance-constrained optimization. *Oper. Res.* **58**(2) 470–485.

Den Hertog, D., P. Stehouwer. 2002. Optimizing color picture tubes by high-cost nonlinear programming. *Eur. J. Oper. Res.* **140**(2) 197–211.

Dupačová, J. 1987. The minimax approach to stochastic programming and illustrative application. *Stochastics* **20** 73–88.

Dupačová, J. 2001. *Encyclopedia of Optimization*. Stochastic Programming: minimax approach. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Epstein, L. G., M. Schneider. 2003. Recursive multiple-priors. *J. Econom. Theory* **113**(1) 1–31.

Epstein, L. G., M. Schneider. 2007. Learning under ambiguity. *Rev. Econom. Stud.* **74**(4) 1275–1303.

Erdoğan, E., G. Iyengar. 2006. Ambiguous chance constrained problems and robust optimization. *Math. Programming* **107**(1–2) 37–61.

Fama, E. F., K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *J. Financial Economics* **33**(1) 3–56.

Gilboa, A. L., D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *J. Mathematical Econom.* **18**(2) 141–153.

Hansen, L. P., T. J. Sargent. 2001. Robust control and model uncertainty. *American Econom. Rev.* **91** 60–66.

Henrion, R., C. Strugarek. 2008. Convexity of chance constraints with independent random variables. *Computational Optim. Appl.* **41** 263–276.

Jager, L., J. A. Weller. 2007. Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053.

Klabjan, D., D. Simchi-Levi, M. Song. 2007. Robust stochastic lot-sizing by means of histograms. Technical report, MIT, Boston, MA.

Miller, B. L., H. M. Wagner. 1965. Chance constrained programming with joint constraints. *Oper. Res.* **13**(6) 930–945.

Myers, R. H., D. C. Montgomery. 1995. *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments*. John Wiley & Sons Inc., New York, NY.

Nemirovski, A., A. Shapiro. 2006. Convex approximations of chance constrained programs. *SIAM J. Optim.* **17**(4) 969–996.

Nemirovski, A., A. Shapiro. 2007. Convex approximations of chance constrained programs. *SIAM J. Optim.* **17**(4) 969–996.

Pardo, L. 2006. *Statistical Inference Based on Divergence Measures.* Chapman & Hall/CRC, Boca Raton, FL.

Prékopa, A. 1970. Efficient robust optimization of metal forming processes using a sequential metamodel based strategy. *In Proceedings of the Princeton Symposium on Mathematical Programming*, Princeton University Press, Princeton, NJ. 113–138.

Prékopa, A. 1973. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum* **34** 335–343.

Prékopa, A. 1974. Programming under probabilistic constraints with a random technology matrix. *Mathematische Operationsforschung und Statistik* **5** 109–116.

Prékopa, A. 1995. *Stochastic Programming.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Shapiro, A., S. Ahmed. 2004. On a class of minimax stochastic programs. *SIAM J. Optim.* **14**(4) 1237–1252.

Shapiro, A., A. J. Kleywegt. 2002. Minimax analysis of stochastic problems. *Optim. Methods and Software* **17** 523–542.

Shi, W. 2011. Design of pre-enhanced cross-docking distribution center under supply uncertainty: RSM robust optimization method. Working Paper, HUST, China.

Wiebenga, J. H., G. Klaseboer, A. H. van den Boogaard. Efficient robust optimization of metal forming processes using a sequential metamodel based strategy. 2011. *8th International Conference and Workshop on Numerical Simulation of 3D Sheet Metal Forming Processes, August 2011*, Korean Society for Technology of Plasticity, Numisheet 2011. 978–985.

Van de Panne, C., W. Popp. 1963. Minimum-cost cattle feed under probabilistic protein constraints. *Management Sci.* **9**(3) 405–430.

Žáčková, J. 1966. On minimax solutions of stochastic linear programs. *Čas. Pěst. Mat.*. 423–430.

Zymler S., D. Kuhn, B. Rüstem. 2012. Distributionally robust joint chance constraints with second-order moment information. Working Paper. *Math. Programming* Forthcoming.

CHAPTER 4

Adjustable Robust Optimization Using Metamodels

4.1 Introduction

In this chapter, we present a methodology that is a novel combination of *robust optimization* (RO) and simulation-based *robust parameter design* (RPD). We mention the major contributions of our research at the end of this section, but before that we extensively describe the fields of RO and simulation-based RPD.

To begin with, the goal of many *experiments* is to estimate the best solution for a given practical problem. Such experiments may be conducted with a physical system (e.g., an airplane model in a wind tunnel) or a mathematical model of a physical system (e.g., a computerized simulation model of an airplane or an inventory management system). These experiments produce data on the outputs or responses for the given inputs or factors. Output may be univariate (a single or scalar response) or multivariate (multiple responses or a vector with responses). The number of inputs may range from a single input to 'many' inputs (e.g., thousands of inputs), but we focus on practical problems with no more than (say) twenty inputs.

Taguchi (1987)—for an update see Myers et al. (2009, pp. 483–485)—distinguishes between the following two types of inputs: (i) Controllable or decision factors (say) d_j $(j = 1, \ldots, k)$, collected in the k-dimensional vector $\mathbf{d} = (d_1, \ldots, d_k)^T$. (ii) Environmental or noise factors (say) e_g $(g = 1, \ldots, c)$, collected in the c-dimensional vector $\mathbf{e} = (e_1, \ldots, e_c)^T$. By definition, the first type of inputs is under the control of the users; e.g., in an inventory system, management controls the order quantity. The second type of inputs is not controlled by the users; e.g., demand in an inventory system. Taguchi emphasizes that the noise factors create variability in the outputs. Consequently, the combination of decision factors that (say) maximizes the expected univariate output may cause the variance of that output to be much larger than a

combination that 'nearly' maximizes the mean output; i.e., a little sacrifice in expected output may save a lot of problems caused by the variability in the output (or as the French proverb states: "the best is the enemy of the good").

Taguchian RPD has been criticized by *statisticians*; see the panel discussion reported in Nair et al. (1992). Their main critique concerns the statistical design and analysis in the Taguchian approach; for details we refer to Taguchi (1987) and Myers et al. (2009, pp. 483–495). An alternative RPD approach is given by Myers et al. (2009, pp. 502–506). Like Myers et al. do, we expose our methodology using the following 'incomplete' second-order polynomial regression metamodel:

$$y(\mathbf{e}, \mathbf{d}) = \beta_0 + \boldsymbol{\beta}^T \mathbf{d} + \mathbf{d}^T \mathbf{B} \mathbf{d} + \boldsymbol{\gamma}^T \mathbf{e} + \mathbf{d}^T \boldsymbol{\Delta} \mathbf{e} + \epsilon \tag{4.1}$$

that approximates the unknown input/output (I/O) function (say) $w = g(\mathbf{e}, \mathbf{d})$ for a single output w of the system, where y denotes the output of the regression metamodel of the simulation model with output w; β_0 the intercept; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ the first-order effects of the decision variables $\mathbf{d}$; $\mathbf{B}$ the $k \times k$ symmetric matrix of second-order effects of $\mathbf{d}$ with on the main diagonal the 'purely quadratic' effects $\beta_{j;j}$ and off this diagonal half the interactions between pairs of decision factors $\beta_{j;j'}/2$ $(j \neq j')$; $\gamma = (\gamma_1, \ldots, \gamma_c)^T$ the first-order effects of the environmental factors $\mathbf{e}$; $\boldsymbol{\Delta} = (\delta_{j;g})$ the $k \times c$ pairwise (two-factor) interactions between $\mathbf{d}$ and $\mathbf{e}$; ϵ the residual with $E(\epsilon) = 0$ if this model has no 'lack of fit' (so it is a valid or adequate approximation of $g(\mathbf{e}, \mathbf{d})$) and with constant variance σ_ϵ^2.

Myers et al. (2009) assume experiments with real systems; whereas we assume experiments with simulation models of real systems. These simulation models may be either deterministic (especially in engineering) or random (stochastic, possibly with discrete events). Deterministic simulation models have noise if a parameter or input variable has a fixed but unknown value; this is called subjective or epistemic uncertainty; see Iman and Helton (2006). Random simulation models also have objective, aleatory or inherent uncertainty; again see Iman and Helton. We focus on random simulation, but shall also discuss deterministic simulation. Simulation analysts use different names for metamodels, such as response surfaces, surrogates, and emulators; see the many references in Kleijnen (2008, p. 8). There are different types of metamodels, but the most popular types are low-order polynomials such as (4.1) and Kriging (Gaussian process) models. These polynomials are nonlinear in the inputs $(\mathbf{e}, \mathbf{d})$ but linear in the regression parameters $(\beta_0, \ldots, \delta_{k;c})$ so the analysis may use classic linear regression models estimated using the least squares (LS) criterion; see Kleijnen (2008, pp. 15–72) and Myers et al. (2009, pp. 13–71). Kriging models are more flexible so they can accurately approximate the true I/O function over bigger

experimental areas; see Kleijnen (2008, pp. 139–156). It is important to point out that methods that shall be proposed in this chapter are not bound by low-order polynomials. Therefore, our methodology is also suitable for Kriging or other classes of metamodels, and these extensions will be mentioned in a later section. Nevertheless, for the sake of exposition, we shall adapt the low-order polynomial notation (4.1) in the remainder of the chapter.

Optimization of systems being simulated—called *simulation optimization* (SO)—is a popular research topic in discrete-event simulation; see Fu (2007), and Fu and Nelson (2003). 'Robust' SO approaches are also discussed in that literature. Examples are Angün (2011) combines Taguchi's worldview with *response surface methodology* (RSM). This RSM is a stepwise optimization heuristic that in the various steps uses local first-order polynomial metamodels and in the final step uses a local second-order polynomial metamodel (obviously, these polynomials are linear regression models). Instead of Taguchi's various criteria, Angün (2011) uses the average value-at-risk (also known as conditional value-at-risk). Miranda and Del Castillo (2011) perform RPD optimization through a well-known SO method; namely, simultaneous perturbation stochastic approximation, which is detailed in Spall (2003). Wiedemann (2010, p. 31) applies Taguchian RPD to an agent-based simulation, ensuring that the mean response meets the target value and the variability around that level is sufficiently small. Some years ago, Al-Aomar (2006) used Taguchi's signal-to-noise ratio and a quality loss function, together with a genetic algorithm with a scalar fitness measure that is a combination of the estimated mean and variance. An older paper is Sanchez (2000), who used Taguchi's worldview with a loss function that incorporates both system's mean and variability, and RSM; the author gave many more references. Robust approaches are discussed—albeit briefly—not only in discrete-event simulation but also in deterministic simulation if that simulation has uncertain environmental variables, which (by definition) are beyond management's control. A recent example is Hu et al. (2012), who propose a robust climate simulation model (called 'DICE') with input parameters that have a multivariate Gaussian distribution with uncertain or 'ambiguous' mean vector and covariance matrix. Another recent example is Dellino et al. (2012), who investigate the well-known economic order quantity model with an uncertain demand rate. Dellino et al. use Taguchi's worldview, but replace his experimental designs (namely, orthogonal arrays) and metamodels (namely, low-order polynomials) by Latin hypercube sampling and Kriging metamodels. Many more applications can be found in engineering. For example, Chan and Ewins (2010) use Taguchi's RPD to manage the vibrations in bladed discs. Urval et al. (2010) use Taguchi's RPD in powder injection moulding of microsystems' components. Delpi-

ano and Sepulveda (2006) also study RPD in engineering. Vanlı and Del Castillo (2009) study *on-line* RPD (ORPD) to account for control variables that are adjusted over time according to the on-line observations of environmental factors. They assume the posterior predictive densities of responses and environmental factors are known; moreover, they assume that the uncertain parameters follow a specific time series model. They propose two Bayesian approaches for ORPD. In both approaches, controllable factors can be adjusted on-line using an expected loss function and the on-line observations. Joseph (2003) proposes RPD with feed-forward control for measurement systems where true values of the responses cannot be observed but estimated. Using the on-line observable environmental factors, the author proposes a control law that periodically updates the online measurement of the system. Dasgupta and Wu (2006) propose an on-line feedback control mechanism that adjusts the observed output error using a controllable factor—called adjustment factor—that is set on-line during production.

Classic RPD assumes that the mean and variance—and sometimes even the probability distribution—of $\mathbf{e}$ are known. The final parameter design may be sensitive to these assumptions. We therefore propose a RO approach that takes the distribution ambiguity into account, and that uses historical data on the environmental inputs. The developments in RO started with Ben-Tal and Nemirovski (1998, 1999) and El-Ghaoui and Lebret (1997). Optimization problems usually have uncertain coefficients in the objective function and the constraints, so the "nominal" optimal solution—i.e., the optimal solution if there would be no uncertainties—may easily violate the constraints for some realizations of the uncertain coefficients. Therefore, it is better to find a "robust" solution, which is immune to the uncertainty in a so-called uncertainty set. The robust reformulation of a given uncertain mathematical optimization problem is called the *robust counterpart* (RC) problem. The mathematical challenge in RO is to find computationally tractable RCs; see Ben-Tal et al. (2009).

As already said in this chapter, we present a methodology that is a combination of RO and simulation-based RPD. Based on Ben-Tal et al. (2013) and Yanıkoğlu and den Hertog (2013), we use experimental data on the simulated system to derive uncertainty sets—to be used in RO—for the unknown probability distribution of the environmental factors. Furthermore, we also use Ben-Tal, den Hertog, and Vial (2014) to convert associated RCs into explicit and computationally tractable optimization problems. Bingham and Nair (2012) point out that "the noise distributions are rarely known, and the choices are often based on convenience". The

advantage of our approach is that it uses the distribution-free structure of RO; i.e., unlike standard approaches we do not make any assumptions about the distribution of the environmental factors such as a Gaussian (or Normal) distribution. Moreover, we develop *adjustable* RPD for those situations in which some or all of the controllable inputs have a 'wait and see' character; i.e., their values can be adjusted after observing the values of some or all of the environmental inputs. Examples are the adjustment of several controllable chemical process parameters after observing environmental inputs such as temperature and humidity, or the adjustment of the replenishment order at time t according to the realized demands in the preceding $t-1$ periods in a multistage inventory system. We develop *adjustable* RO for such situations, and show that the corresponding robust counterpart problems can again be reformulated as tractable optimization problems.

The *major contributions* of our research can be summarized as follows: (i) We propose a RO methodology for a class of RPD problems where the distributional parameters are unknown but historical data on the uncertain parameters are available. (ii) Our method is suitable for different classes of metamodels, e.g., higher-order polynomials, Kriging, and radial basis functions. (iii) We propose adjustable robust approach for RPD. Unlike classic adjustable RO techniques, our adjustable robust reformulations are tractable (i.e., when the number of uncertain parameters is fixed) even for nonlinear decision rules. (iv) We propose tractable RC formulations of uncertain optimization problems that have quadratic terms in the uncertain parameters. Compared with other studies in the literature, our formulations can handle more general classes of uncertainty sets, and lead to easier tractable formulations. (v) Last but not least, we introduce adjustable integer decision variables in the context of RO, and propose a specific decision rule for such variables. Finally, the *limitation* our method is that it is suitable for low dimensional uncertainties.

Related to our research, Stinstra and den Hertog (2008) and Bertsimas et al. (2010) also propose RO methods for different classes of SO problems. More precisely, Stinstra and den Hertog (2008) propose RO methods for three types of errors, namely, simulation-model, metamodel, and implementation errors, that may arise at different stages of a simulation-based optimization approach. The authors consider box and ellipsoidal uncertainties for the associated errors. Bertsimas et al. (2010) on the other hand propose a RO method for unconstrained simulation-based problems with non-convex cost functions. This method is suitable for metamodels that are often used in practice such as Kriging, and it can be generalized to both implementation errors and parameter uncertainties.

The remainder of the chapter is organized as follows. §4.2 summarizes Taguchi's world view, and alternative RPD approaches in SO. §4.3 proposes our RO approach accounting for distribution ambiguity. §4.4 develops adjustable RO approaches. §4.5 presents numerical examples. §4.6 summarizes our conclusions, and indicates a future research topic.

4.2 Robust Parameter Design

In this section, we first present the Taguchian RPD approach, and then other popular RPD approaches. To estimate the regression parameters in (4.1) we need an experiment with the underlying system: Design of experiments (DoE) uses *coded*—also called *standardized* or *scaled*—values (say) x_j for the factors. So the experiment consists of (say) n combinations of the coded factors $\mathbf{x}$, which correspond with $\mathbf{d}$ and $\mathbf{e}$ in (4.1). Coding is further discussed in Kleijnen (2008, p. 29) and Myers et al. (2009, p. 78).

We reformulate the Taguchian model (4.1) as the following *linear regression model*:

$$y = \boldsymbol{\zeta}^T \mathbf{x} + \epsilon \tag{4.2}$$

with the ℓ-dimensional vector of regression parameters or coefficients $\boldsymbol{\zeta} = (\beta_0, \ldots, \delta_{k;c})^T$ and the corresponding vector of regression explanatory variables $\mathbf{x}$ defined in the obvious way; e.g., the explanatory variable corresponding with the interaction effect $\beta_{1;2}$ is $d_1 d_2$. Then (4.2) leads to the LS estimator:

$$\widehat{\boldsymbol{\zeta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \tag{4.3}$$

where $\mathbf{X}$ is the $n \times \ell$ matrix of explanatory variables with n denoting the number of scenarios (combinations of control and environmental factors) determined by the DoE that are actually observed (in a real or simulated system); $\mathbf{y}$ is the vector with the n observed outputs. The covariance matrix of the estimator (4.3) is

$$\text{Cov}(\widehat{\boldsymbol{\zeta}}) = \sigma_\epsilon^2 (\mathbf{X}^T \mathbf{X})^{-1} \tag{4.4}$$

where σ_ϵ^2 was defined in (4.1). Hence, the variance σ_ϵ^2 may be estimated by the mean squared residuals:

$$\hat{\sigma}_\epsilon^2 = \frac{(\widehat{\mathbf{y}} - \mathbf{y})^T (\widehat{\mathbf{y}} - \mathbf{y})}{n - \ell} \tag{4.5}$$

where $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\zeta}}$; see (4.2) and (4.3).

We denote the expected values of the environmental factors through $\mathrm{E}(\mathbf{e}) = \boldsymbol{\mu}_{\mathbf{e}}$. We allow dependence between the environmental factors so the covariance matrix is $\mathrm{Cov}(\mathbf{e}) = \boldsymbol{\Omega}_{\mathbf{e}}$. Analogous to Myers et al. (2009, pp. 504–506), we derive that the metamodel (4.1) implies that the regression predictor for the mean $\mathrm{E}(y)$ is

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] = \beta_0 + \boldsymbol{\beta}^T \mathbf{d} + \mathbf{d}^T \mathbf{B}\mathbf{d} + \boldsymbol{\gamma}^T \boldsymbol{\mu}_{\mathbf{e}} + \mathbf{d}^T \boldsymbol{\Delta}\mu_{\mathbf{e}} \tag{4.6}$$

and the regression predictor for the variance $\mathrm{Var}(y)$ is

$$\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] = (\boldsymbol{\gamma}^T + \mathbf{d}\boldsymbol{\Delta})\boldsymbol{\Omega}_{\mathbf{e}}(\boldsymbol{\gamma} + \boldsymbol{\Delta}^T \mathbf{d}) + \sigma_\epsilon^2 \tag{4.7}$$

where $(\boldsymbol{\gamma} + \boldsymbol{\Delta}^T \mathbf{d}) = (\partial y/\partial e_1, \ldots, \partial y/\partial e_c)^T$ is the gradient with respect to the environmental factors. Obviously, the larger the gradient's elements are, the larger the variance of the predicted output is. Furthermore, if $\boldsymbol{\Delta} = \mathbf{0}$ (no control-by-noise interactions), then $\mathrm{Var}(y)$ cannot be controlled through the decision variables $\mathbf{d}$. Note the difference between the predicted variance, $\mathrm{Var}(y)$, and the variance of the predictor, $\mathrm{Var}(\widehat{y})$ with $\widehat{y} = \mathbf{x}^T \widehat{\boldsymbol{\zeta}}$. Obviously, the mean vector and the covariance matrix completely define a multi-variate Gaussian distribution.

Taguchi focuses his analysis on the *signal-to-noise ratios* (SNRs), which depend on $\mathrm{E}(y)/\sqrt{\mathrm{Var}(y)}$ (the standard deviation $\sqrt{\mathrm{Var}(y)}$ has the same scale as the mean $\mathrm{E}(y)$); see Myers et al. (2009, pp. 486–488). The precise definitions of these SNRs vary with the following goals of RPD: (i) 'The smaller, the better': minimize the response. (ii) 'The larger, the better': maximize the response. (iii) 'The target is best': realize a target value (say) T for the response.

We do not further dwell on Taguchi's SNRs, because we think there are better formulations of the various goals of RPD; Myers et al. (2009, pp. 488–495) also question the utility of SNRs. We use the following optimization problem formulation, also formulated by Myers et al. (2009, p. 506):

$$\min_{\mathbf{d}} \ \mathrm{Var}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] \ \text{ s.t. } \ \mathrm{E}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] \leq T. \tag{4.8}$$

We may also consider the optimization problem given by Dellino et al. (2012); namely,

$$\min_{\mathbf{d}} \ \mathrm{E}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] \ \text{ s.t. } \ \mathrm{Var}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] \leq T, \tag{4.9}$$

or the following variant:

$$\min_{\mathbf{d}} \ \mathrm{E}_{\mathbf{e}}[(y(\mathbf{e}, \mathbf{d}) - T)^2]. \tag{4.10}$$

Remark 8 *The optimal 'coded' controllable factors* $\mathbf{d}^*$*—that minimize the process mean or variance in (4.8), (4.9) or (4.10)—must lie in* $-\mathbf{1} \leq \mathbf{d} \leq \mathbf{1}$*, where* $\mathbf{1}$ *denotes the all ones vector. Otherwise, the simulation or physical experiment must be rerun with a larger experimental region for the original (non-coded) input factors to obtain a new response model* $y(\mathbf{e}, \mathbf{d})$ *satisfying the associated requirement for the given RPD problem.*

In the rest of this chapter we focus on (4.9), since (4.8), (4.10) or other variants can be treated analogously.

4.3 Robust Optimization with Unknown Distributions

In this section, we derive the robust reformulations of the class of optimization problems presented in §4.2. We assume that data on the environmental factors $\mathbf{e}$ is available or can be obtained via simulation.

4.3.1 Uncertainty Sets

Instead of relying on the normal distribution, RO derives an *uncertainty set* for the unknown density function of the noise factors. There are several RO approaches, but we follow Ben-Tal et al. (2013) and Yanıkoğlu and den Hertog (2013), who develop RO accounting for *historical data* on the noise factors. They do not use a specific distribution to this data; instead, they use the more general concept of *ϕ-divergence*—as follows.

Given N historical observations on the noise factors $\mathbf{e}$, we construct m cells such that the number of observations o_i in cell i $(i = 1, \ldots, m)$ is at least five:

$\sum_{i=1}^{m} o_i = N$ such that $\forall i : o_i \geq 5$.

The historical data on $\mathbf{e}$ give the frequencies $\mathbf{q} = [q_1, \ldots, q_m]^T$, where q_i is the observed frequency in cell i so

$$q_i = \frac{o_i}{N}.$$

We consider $\mathbf{p} = [p_1, \ldots, p_m]^T$ as the unknown true probability vector of $\mathbf{e}$ when its support is discretized into m cells; hence $\mathbf{q}$ is an approximation of $\mathbf{p}$. The ϕ-divergence measure is

$$I_\phi(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right)$$

where $\phi(.)$ satisfy certain mathematical requirements such as $\phi(1) = 0$, $\phi(a/0) := a \lim_{t\to\infty} \phi(t)/t$ for $a > 0$, and $\phi(0/0) = 0$; for details on ϕ-divergence we refer to Pardo (2006). It can be proven that the test statistic

$$\frac{2N}{\phi''(1)} I_\phi(\mathbf{p}, \mathbf{q})$$

is asymptotically distributed as a chi-squared random variable with $(m-1)$ degrees of freedom. So an asymptotic $(1-\alpha)$-confidence set for $\mathbf{p}$ is

$$I_\phi(\mathbf{p}, \mathbf{q}) \leq \rho \text{ with } \rho = \rho(N, m, \alpha) = \frac{\phi''(1)}{2N} \chi^2_{m-1;1-\alpha}. \tag{4.11}$$

Using (4.11), Ben-Tal et al. (2013) derive the following *uncertainty set* U for the unknown probability vector $\mathbf{p}$:

$$U = \{\mathbf{p} \in \mathbb{R}^m | \mathbf{p} \geq 0, \sum_{i=1}^{m} p_i = 1, I_\phi(\mathbf{p}, \mathbf{q}) \leq \rho\}. \tag{4.12}$$

Table 4.1 – ϕ-Divergence Examples

Divergence	$\phi(t), t > 0$	$I_\phi(\mathbf{p}, \mathbf{q})$	$\phi^*(s)$
Kullback-Leibler	$t \log t$	$\sum_i p_i \log\left(\frac{p_i}{q_i}\right)$	e^{s-1}
Burg entropy	$-\log t$	$\sum_i q_i \log\left(\frac{p_i}{q_i}\right)$	$-1 - \log(-s), s \leq 0$
χ^2-distance	$\frac{1}{t}(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{p_i}$	$2 - 2\sqrt{1-s}, s \leq 1$
Pearson χ^2-distance	$(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{q_i}$	$s + s^2/4, s \geq -2$ $-1, s < -2$
Hellinger distance	$(1-\sqrt{t})^2$	$\sum_i (\sqrt{p_i} - \sqrt{q_i})^2$	$\frac{s}{1-s}, s \leq 1$

Table 4.1 taken from Ben-Tal et al. (2013) presents common choices of the ϕ-divergence function; $\phi^*(s) := \sup_{t\geq 0}\{st - \phi(t)\}$ denotes the conjugate of a ϕ-divergence distance that will be used for the derivations of RCs in the next section. In various examples, we shall use the χ^2-distance in this chapter.

4.3.2 Robust Counterparts of Mean and Variance

In this section we derive the robust reformulation of the optimization problem (4.9), given in §4.2. To point out, the RCs of other classes of problems based on mean and/or variance, e.g., (4.8) and (4.10), may be derived by slightly modifying the

derivations in this section.

We represent each cell i by its center point $\mathbf{e}^i$; e.g., e^i_g is the g^{th} $(g = 1, \ldots, c)$ coordinate of cell i's center point. To apply the ϕ-divergence in (4.12), we discretize $y(\mathbf{e}, \mathbf{d})$ by replacing $\mathbf{e}$ by $\mathbf{e}^i$. This substitution gives

$$y_i(\mathbf{d}) = y(\mathbf{e}^i, \mathbf{d}) = \beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + (\boldsymbol{\gamma}^T + \mathbf{d}^T\Delta)\mathbf{e}^i. \tag{4.13}$$

Consequently, the mean of $y(\mathbf{e}, \mathbf{d})$ becomes approximated by

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] = \sum_{i\in V} y_i(\mathbf{d})p_i = \sum_{i\in V}[\beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + (\gamma^T + \mathbf{d}^T\Delta)\mathbf{e}^i]p_i \tag{4.14}$$

where p_i denotes the probability of $\mathbf{e}$ falling into cell i. Note that $\mathbf{p}$ is in the uncertainty set U given by (4.12) and the empirical estimate $\mathbf{q}$ of $\mathbf{p}$ is obtained using the data on $\mathbf{e}$. The limitation of the approach is that it requires excessive data points in high dimensional uncertainties.

Next we define

$$\psi_i(\mathbf{d}) := (\boldsymbol{\gamma}^T + \mathbf{d}^T\Delta)\mathbf{e}^i, \tag{4.15}$$

and the variance of $y(\mathbf{e}, \mathbf{d})$ becomes approximated by

$$\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})] = \sum_{i\in V} \psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V} \psi_i(\mathbf{d})p_i\right]^2. \tag{4.16}$$

Eventually, the approximate *robust reformulation* of (4.9) is the following *semi-infinite* optimization problem:

$$\begin{aligned}(\mathbf{SI1})\ \min_{\mathbf{d}}\ \max_{\mathbf{p}\in U} & \sum_{i\in V}[\beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + (\gamma^T + \mathbf{d}^T\Delta)\mathbf{e}^i]p_i \\ \text{s.t.} & \sum_{i\in V} \psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V} \psi_i(\mathbf{d})p_i\right]^2 \le T \qquad \forall \mathbf{p} \in U, \end{aligned} \tag{4.17}$$

where $U = \{\mathbf{p} \in \mathbb{R}^m | \mathbf{p} \ge 0, \sum_{i=1}^m p_i = 1, I_\phi(\mathbf{p}, \mathbf{q}) \le \rho\}$. (SI1) is a difficult optimization problem that has infinitely many constraints (see $\forall \mathbf{p} \in U$), and includes quadratic terms in $\mathbf{p}$. Ben-Tal et al. (2009, p. 382) propose a tractable RC of a linear optimization problem with uncertain parameters that appear quadratically, and an ellipsoidal uncertainty. The resulting formulation is a semidefinite programming (SDP) problem; see also Remark 9 below. The following theorem provides tractable RC reformulations of (SI1) for more general 'ϕ-divergence' uncertainty sets.

Theorem 3 *The vector* $\mathbf{d}$ *solves (SI1) if and only if* $\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\eta}$*, and* z *solve the following RC problem:*

$$\begin{aligned}(\mathbf{RC1}) \min_{\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\eta},z} \; & \beta_0 + \boldsymbol{\beta}^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d} + \lambda_1 + \rho\eta_1 + \eta_1 \sum_{i\in V} q_i \phi^* \left(\frac{\psi_i(\mathbf{d}) - \lambda_1}{\eta_1}\right) \\ s.t. \; & \lambda_2 + \rho\eta_2 + \eta_2 \sum_{i\in V} q_i \phi^* \left(\frac{(\psi_i(\mathbf{d}) + z)^2 - \lambda_2}{\eta_2}\right) \leq T \\ & \eta_1, \eta_2 \geq 0\end{aligned} \tag{4.18}$$

where ρ *is given by (4.11),* $\phi^*(s) := \sup_{t\geq 0}\{st - \phi(t)\}$ *denotes the conjugate of* $\phi(.)$*,* $V = \{1, ..., m\}$ *is the set of cell indices in the uncertainty set* U*,* q_i *is the data frequency in cell* $i \in V$ *using the historical data on* $\mathbf{e}$*, and* $\boldsymbol{\lambda}$*,* $\boldsymbol{\eta}$*, and* z *are additional variables.*

Proof. Using Yanıkoğlu and den Hertog (2013, Theorem 1), we can easily derive the explicit RC of the objective function of (SI1) that is linear in $\mathbf{p} \in U$. Next we consider the 'more difficult' variance constraint (4.17), which is quadratic in $\mathbf{p}$. In the following parts of the proof we use Ben-Tal, den Hertog, and Vial (2014) to account for the nonlinear uncertainty in the constraint. Using a linear transformation, we reformulate (4.17) as

$$\max_{\mathbf{a}\in\widehat{U}} g(\mathbf{a}) \leq T, \tag{4.19}$$

where $\widehat{U} := \{\mathbf{a} : \mathbf{a} = \mathbf{A}\mathbf{p}, \mathbf{p} \in U\}$, $\mathbf{A}^T = [\psi^2(\mathbf{d}), \psi(\mathbf{d})]$ and $g(\mathbf{a}) = a_1 - a_2^2$. Using the indicator function

$$\delta(\mathbf{a}|\widehat{U}) := \begin{cases} 0, & \mathbf{a} \in \widehat{U} \\ +\infty, & \text{elsewhere,} \end{cases}$$

we reformulate (4.19) as

$$\max_{\mathbf{a}\in\mathbb{R}^2}\{g(\mathbf{a}) - \delta(\mathbf{a}|\widehat{U})\} \leq T. \tag{4.20}$$

The Fenchel dual of (4.20)—for details see Rockafellar (1970, pp. 327–341)— is equivalent to

$$\min_{\mathbf{v}\in\mathbb{R}^2}\{\delta^*(\mathbf{v}|\widehat{U}) - g_*(\mathbf{v})\} \leq T \tag{4.21}$$

where $\mathbf{v}$ denotes the dual variable, and $\delta^*(\mathbf{v}|\widehat{U}) := \sup_{\mathbf{p}\in U}\{\mathbf{a}^T\mathbf{v}|\mathbf{a} = \mathbf{A}\mathbf{p}\}$ and $g_*(\mathbf{v}) := \inf_{\mathbf{a}\in\mathbb{R}^2}\{\mathbf{a}^T\mathbf{v} - g(\mathbf{a})\}$ denote the convex and concave conjugates of the functions δ and g, respectively. Going from (4.20) to (4.21) is justified since the

intersection of the relative interiors of the domains of $g(.)$ and $\delta(.|\widehat{U})$ is non-empty, since $\mathbf{a} = \mathbf{Aq}$ is always in the relative interiors of both domains. Moreover it is easy to show that $\delta^*(\mathbf{v}|\widehat{U}) = \delta^*(\mathbf{A}^T\mathbf{v}|U)$. Then we delete the minimization in (4.21) because the constraint has the $\leq$ operator, and the RC reformulation of (4.17) becomes

$$\delta^*(\mathbf{A}^T\mathbf{v}|U) - g_*(\mathbf{v}) \leq T.$$

Now we derive the complete formulas of the conjugate functions δ^* and g_*. If $\mathbf{v}^T = [w, z]$, then the concave conjugate of g is equivalent to

$$g_*(\mathbf{v}) = \inf_{\mathbf{a}\in\mathbb{R}^2}\{a_1 w + a_2 z - g(\mathbf{a})\} = \begin{cases} -z^2/4, & w = 1 \\ -\infty, & \text{elsewhere.} \end{cases}$$

Using Theorem 1 in Yanıkoğlu and den Hertog (2013) once more, the convex conjugate of δ is equivalent to

$$\delta^*(\mathbf{A}^T\mathbf{v}|U) = \inf_{\lambda,\eta_2\geq 0}\left\{\rho\eta_2 + \lambda + \eta_2\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d}) + \psi_i(\mathbf{d})z - \lambda}{\eta_2}\right)\right\}.$$

Thus the RC reformulation of (4.17) becomes

$$\lambda + \rho\eta_2 + \eta_2\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d}) + \psi_i(\mathbf{d})z - \lambda}{\eta_2}\right) + \frac{z^2}{4} \leq T, \eta_2 \geq 0. \tag{4.22}$$

Substituting $\lambda_2 = \lambda + z^2/4$ into (4.22) gives the final RC reformulation (RC1) ■

Remark 9 *An ellipsoidal uncertainty set is a special case of the ϕ-divergence uncertainty set* (4.12) *when the Pearson chi-squared distance is used as the ϕ-divergence. Moreover, we can reformulate (RC1) as a second order cone problem (SOCP) for the associated distance measure. Notice that SOCP is an 'easier' formulation of the problem compared with the SDP by Ben-Tal et al. (2009).*

Remark 10 *Ben-Tal et al. (2014) also propose an RC for the variance uncertainty, however the associated RC introduces additional non-convexity. We overcome this difficulty by using the substitution $\lambda_2 = \lambda + z^2/4$ in the proof of Theorem 1.*

We now discuss the 'general' computational tractability of (RC1). First, $\phi^*(h(\mathbf{d}, \lambda_2, z))$ is convex, since the convex conjugate $\phi^*(s) := \sup_{t\geq 0}\{st - \phi(t)\}$ is non-decreasing in s, and $h(\mathbf{d}, \lambda_2, z) = (\psi_i(\mathbf{d}) + z)^2 - \lambda_2$ is convex in $\mathbf{d}, z$, and λ_2. It is easy to show that $\eta_2\phi^*(\cdot/\eta_2)$ is convex, since the perspective of a convex function is always convex. Eventually, the convexity of the perspective implies that (4.18) is convex. On the other hand, the objective function of (RC1) is not necessarily convex, since y is non-convex in $\mathbf{d}$ unless $\mathbf{B}$ is a positive semidefinite (PSD) matrix. Nevertheless, (RC1) does not introduce additional non-convexity into the general optimization problem (4.9).

4.3.3 Alternative Metamodels and Risk Measures

In this subsection, we focus on extensions of our method. §4.3.3.1 presents a generalization of our method for other metamodels besides (4.2), and §4.3.3.2 presents an extension of our method to SNRs. Finally, §4.3.3.3 shows how to apply our method to tail-risk measures.

4.3.3.1 Alternative Metamodels

In this chapter we focus on the low-order polynomial (4.1), since most of the literature and real-life applications use low-order polynomials to approximate the I/O function of the underlying simulation or physical experiment. However, our methodology can also be used to other metamodel types such as higher-order polynomials, Kriging, and radial basis functions. More precisely, consider

$$y(\mathbf{d}, \mathbf{e}) = f(\mathbf{d}) + \psi(\mathbf{d}, \mathbf{e})$$

where $f(\mathbf{d})$ is the part that affects only the mean of the response (e.g., it is $f(\mathbf{d}) = \beta_0 + \beta^T\mathbf{d} + \mathbf{d}^T\mathbf{B}\mathbf{d}$ in (4.1)), and $\psi(\mathbf{d}, \mathbf{e})$ is the part that affects the response variance (e.g., it is $\psi(\mathbf{d}, \mathbf{e}) = \boldsymbol{\gamma}^T\mathbf{e} + \mathbf{d}^T\boldsymbol{\Delta}\mathbf{e}$ in (4.1)). We then reformulate the RC in Theorem 1 as

$$\begin{aligned} \min_{\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\eta},z} \; & f(\mathbf{d}) + \lambda_1 + \rho\eta_1 + \eta_1 \sum_{i\in V} q_i \phi^* \left(\frac{\psi_i(\mathbf{d}) - \lambda_1}{\eta_1} \right) \\ \text{s.t.} \; & \lambda_2 + \rho\eta_2 + \eta_2 \sum_{i\in V} q_i \phi^* \left(\frac{(\psi_i(\mathbf{d}) + z)^2 - \lambda_2}{\eta_2} \right) \leq T \\ & \eta_1, \eta_2 \geq 0. \end{aligned}$$

The complexity of $f(\mathbf{d})$ and $\psi_i(\mathbf{d})$ $(= \psi(\mathbf{d}, \mathbf{e}^\mathbf{i}))$ determines the complexity of the RC; the problem can be non-convex depending on $f(\mathbf{d})$ and $\psi_i(\mathbf{d})$. Again our robust reformulation does introduces additional variables, but does not introduce additional non-convexity.

4.3.3.2 Signal-to-Noise Ratios

SNRs are performance criteria used in many areas including engineering, chemistry, and physics. As we have mentioned earlier in §4.2, Taguchi (1987) focuses on three performance measures in his SNRs. Our method can be applied for any given function, including the first two measures; namely, $(y(\mathbf{e}, \mathbf{d}) - 0)^2$ and $(1/y(\mathbf{e}, \mathbf{d}))^2$. Therefore, the robust reformulations of the given problems are special cases of Theorem 1, when we have no constraint on the variance. The third measure involves a true SNR that is in line with the following expression:

$$\max_{\mathbf{d}} \frac{\mathrm{E}_\mathbf{e}[y(\mathbf{e}, \mathbf{d})]}{\sqrt{\mathrm{Var}_\mathbf{e}[y(\mathbf{e}, \mathbf{d})]}}, \tag{4.23}$$

which we reformulate as

$$\max_{\mathbf{d},w} \; w \quad \text{s.t.} \; -\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] + w\sqrt{\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]} \leq 0. \tag{4.24}$$

by using an epigraphic reformulation and the additional variable $w \in \mathbb{R}$. Like in §4.3.3.1, we define the response model through $y(\mathbf{e},\mathbf{d}) = f(\mathbf{d}) + \psi(\mathbf{e},\mathbf{d})$. Using the associated notation, the general RC of (4.24) after discretization is given by

$$(\mathbf{SI2}) \max_{\mathbf{d},\lambda,\eta,w,\mathbf{v}} \; w$$
$$\text{s.t.} \; -f(\mathbf{d}) - \sum_{i\in V}\psi(\mathbf{d})p_i + w\sqrt{\sum_{i\in V}\psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2} \leq 0 \quad \forall \mathbf{p} \in U.$$

The left-hand side of the inequality in (SI2) is concave in $\mathbf{p}$ when $w \geq 0$. The next theorem provides the tractable RC of (SI2).

Theorem 4 *The vector* $\mathbf{d}$ *and* w *solve (SI2) if and only if* $\mathbf{d}, \lambda, \eta, w,$ *and* $\mathbf{v} = [v_1, v_2]$ *solve the following RC constraints:*

$$(\mathbf{RC2}) \max_{\mathbf{d},\lambda,\eta,w,\mathbf{v}} \; w$$
$$\begin{aligned} s.t. \; & -f(\mathbf{d}) + \lambda + \rho\eta + \frac{w^2}{4v_1}... \\ & + \eta\sum_{i\in V} q_i\phi^*\left(\eta^{-1}\left[\left(\psi_i(\mathbf{d})\sqrt{v_1} + v_2\right)^2 - \psi_i(\mathbf{d}) - \lambda_1\right]\right) \leq 0, \\ & \eta \geq 0, w \geq 0. \end{aligned}$$

Proof. We reformulate the semi-infinite problem (SI2) as

$$\max_{\mathbf{p}\in U}\left\{-f(\mathbf{d}) - \sum_{i\in V}\psi(\mathbf{d})p_i + w\sqrt{\sum_{i\in V}\psi_i^2(\mathbf{d})p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2}\right\} \leq T. \tag{4.25}$$

Analogous to the proof of Theorem 1, we reformulate (4.25) as

$$-f(\mathbf{d}) + \max_{\mathbf{a}\in\widehat{U}} g(\mathbf{a}) \leq T, \tag{4.26}$$

where $\widehat{U} := \{\mathbf{a} : \mathbf{a} = \mathbf{A}\mathbf{p}, \mathbf{p} \in U\}$, $\mathbf{A}^T = [\psi^2(\mathbf{d}), \psi(\mathbf{d})]$ and $g(\mathbf{a}) = -a_2 + w\sqrt{a_1 - a_2^2}$. Using the indicator function

$$\delta(\mathbf{a}|\widehat{U}) := \begin{cases} 0, & \mathbf{a} \in \widehat{U} \\ +\infty, & \text{elsewhere,} \end{cases}$$

we reformulate (4.26) as

$$-f(\mathbf{d}) + \max_{\mathbf{a}\in\mathbb{R}^2}\{g(\mathbf{a}) - \delta(\mathbf{a}|\widehat{U})\} \leq T.$$

Deleting the minimization in the Fenchel dual of $\max_{\mathbf{a}\in\mathbb{R}^2}\{g(\mathbf{a}) - \delta(\mathbf{a}|\widehat{U})\}$ as in Theorem 1, the RC is equivalent to

$$-f(\mathbf{d}) + \delta^*(\mathbf{v}|\widehat{U}) - g_*(\mathbf{v}) \leq T \tag{4.27}$$

where the concave conjugate is

$$g_*(\mathbf{v}) = \inf_{\mathbf{a}\in\mathbb{R}^2}\{a_1v_1 + a_2v_2' - g(\mathbf{a})\} = \begin{cases} -[w^2 + (1+v_2')^2]/4v_1, & v_1 > 0 \\ -\infty, & \text{elsewhere,} \end{cases}$$

and $\mathbf{v} = [v_1; v_2']$ denotes the additional dual variables. The convex conjugate of δ is equivalent to

$$\delta^*(\mathbf{A}^T\mathbf{v}|U) = \inf_{\lambda',\eta\geq 0}\left\{\rho\eta + \lambda' + \eta\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d})v_1 + \psi_i(\mathbf{d})v_2' - \lambda'}{\eta}\right)\right\}$$

where η and λ' are the additional Lagrangian dual variables. Thus (4.27) becomes

$$-f(\mathbf{d}) + \rho\eta + \lambda' + \eta\sum_{i\in V} q_i\phi^*\left(\frac{\psi_i^2(\mathbf{d})v_1 + \psi_i(\mathbf{d})v_2' - \lambda'}{\eta}\right) + \frac{w^2 + (1+v_2')^2}{4v_1} \leq T$$

$$\eta \geq 0, w \geq 0.$$

Using the substitutions $\lambda = \lambda' + (1+v_2)^2/4v_1$ and $v_2 = v_2' + 1/2\sqrt{v_1}$, the final RC becomes RC2 ■

Notice that when v_1 is fixed, RC2 does not introduce extra non-convexity. We can find the optimal v_1 by solving the problem for various values of v_1.

4.3.3.3 Using Standard Deviation as Risk Measure

Standard deviation risk measures are used in finance and engineering to quantify the worst-case risk. In this subsection, we examine the robust reformulation of such constraint given as

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] + k\sqrt{\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]} \leq T$$

where $k \geq 0$. Analogous to the previous subsection, the general RC is given by

$$\textbf{(SI3)}\ f(\mathbf{d}) + \sum_{i\in V}\psi(\mathbf{d})p_i + k\sqrt{\sum_{i\in V}\psi_i(\mathbf{d})^2p_i - \left[\sum_{i\in V}\psi_i(\mathbf{d})p_i\right]^2} \leq T \qquad \forall\mathbf{p}\in U.$$

Corollary 1 *The vector* $\mathbf{d}$ *solves constraint (SI3) if and only if* $\mathbf{d}, \lambda, \eta$, *and* $\mathbf{v} = [v_1, v_2]$ *solve the following RC constraints:*

$$f(\mathbf{d}) + \lambda + \rho\eta + \frac{k^2}{4v_1} + \eta \sum_{i \in V} q_i \phi^* \left((\psi_i(\mathbf{d})\sqrt{v_1} + v_2)^2 + \psi_i(\mathbf{d}) - \lambda_1 \right) \leq T,$$

$$\eta \geq 0.$$

Proof. The proof follows from Theorem 2, when $g(\mathbf{a}) = a_2 + k\sqrt{a_1 - a_2^2}$ ■

4.4 Adjustable Robust Optimization

In the preceding sections the controllable factors $\mathbf{d}$ are 'here and now' decisions; i.e., decisions on $\mathbf{d}$ must be made before $\mathbf{e}$ are realized, and hence $\mathbf{d}$ do not depend on the actual values of $\mathbf{e}$. In practice, a part of the controllable factors can often be adjusted after observing the actual values of $\mathbf{e}$. For example, in a multi-stage inventory system with uncertain demand, the decisions on the replenishment orders are made one-at-a-time, and the replenishment order t is placed when the actual demands in periods 1 through $t-1$ are known. Hence it is realistic to allow the replenishment order for period t to be adjusted according to the demands in the preceding periods, even though the upcoming demands remain uncertain. The adjustable factors are called 'wait and see' decisions. Often here-and-now and wait-and-see decisions appear together in the same problem setting.

To model this situation, Ben-Tal et al. (2009, Chapter 14) reformulate adjustable factors as functions of the uncertain parameters as follows:

$$d_j = D_j(\mathbf{x_j}, \mathbf{P_j e}) \quad j \in \{1, \ldots, n\}, \tag{4.28}$$

where $D_j(.)$ are the so-called decision rules that define the class of functions (e.g., affine functions), $\mathbf{x_j}$ is the vector of coefficient variables to be optimized for the associated function class, and $\mathbf{P_j}$ is the information-base matrix; e.g., if $\mathbf{P_j}$ is a zero matrix, then $\mathbf{d}$ and $\mathbf{e}$ become functionally independent and we are back to here-and-now decisions; if $\mathbf{P_j}$ is a unit matrix, then we allow d_j to depend on all components of $\mathbf{e}$. In addition, d_j can depend on a portion of the observed data on $\mathbf{e}$; e.g., in the multi-stage inventory example, $\mathbf{P_j}$ has the value 1 in the first $j-1$ diagonal elements and zero elsewhere.

To obtain the *adjustable robust counterpart* (ARC) of the mean-variance problem (SI1), we replace $\mathbf{d}$ by $\mathbf{D}(\mathbf{X}, \mathbf{Pe}) = [D_1(\mathbf{x_1}, \mathbf{P_1 e}), \ldots, D_n(\mathbf{x_n}, \mathbf{P_n e})]^T$ in the general

RC:

$$\min_{\mathbf{X}} \max_{\mathbf{p}\in U} \sum_{i\in V}[\beta_0 + \boldsymbol{\beta}^T \mathbf{D}(\mathbf{X}, \mathbf{Pe^i}) + \mathbf{X}, \mathbf{Pe^i})^T \mathbf{B}\, \mathbf{D}(\mathbf{X}, \mathbf{Pe^i}) + (\boldsymbol{\gamma}^T + \mathbf{D}(\mathbf{X}, \mathbf{Pe^i})^T \Delta)\mathbf{e}^i]p_i$$

$$\text{s.t.} \sum_{i\in V} y_i \left(\mathbf{D}(\mathbf{X}, \mathbf{Pe^i})\right)^2 p_i - \left[\sum_{i\in V} y_i \left(\mathbf{D}(\mathbf{X}, \mathbf{Pe^i})\right) p_i\right]^2 \leq T \;\; \forall \mathbf{p} \in U.$$

The tractable ARC of the above problem results from Theorem 1 in §4.3.2. The adjustable reformulation has the following advantages as we shall detail below. First, the optimal solution of ARC is less conservative than that of the non-adjustable RC, since it is flexible in adjusting the robust optimal decisions according to revealed data. Second, our approach is tractable even for nonlinear decision rules because we translate the uncertainty from $\mathbf{e}$ to $\mathbf{p}$. To point out, in classic *adjustable robust optimization* (ARO) tractability is generally scarce for non-linear decision rules; Ben-Tal et al. (2009, pp. 382-387) show that the explicit ARC for a quadratic decision rule is derivable only for an ellipsoidal uncertainty set, and the resulting ARC is an SDP problem. Third, similar to Bertsimas and Caramanis (2010) and Hadjiyiannis et al. (2011), we propose a decision rule that enables modeling integer adjustable decision variables in multistage optimization problems.

To present the associated decision rules we use the following illustrative example:

$$\min_{\mathbf{d}} \; \mathrm{E}_{\mathbf{e}} \left[(1 + 5d_1 + 5d_2 + e_1 - e_2)^2 + (1 + 5d_1 + 10d_2 + e_1 + e_2)^2\right] \tag{4.29}$$

where all factors are coded. We divide the support of $\mathbf{e}$—namely, the unit box in two dimension $[-1, 1]^2$—into four cells of equal size. Hence the center points of the cells are $\{\mathbf{e^1}, \mathbf{e^2}, \mathbf{e^3}, \mathbf{e^4}\}$ =$\{$(-0.5, -0.5), (-0.5, 0.5), (0.5, -0.5), (0.5, 0.5)$\}$). So the nominal problem after discretization is

$$\min_{\mathbf{d}} \sum_{i=1}^{4} \left[\left(1 + 5d_1 + 5d_2 + e_1^i - e_2^i\right)^2 + \left(1 + 5d_1 + 10d_2 + e_1^i + e_2^i\right)^2\right] q_i. \tag{4.30}$$

Suppose the observed data are $\mathbf{q} = [0.4, 0.3, 0.2, 0.1]^T$. It is easy to derive the optimal solution of this *nominal* problem: (d_1, d_2) =(-0.08, -0.08). The uncertainty set for the RC is given by

$$\mathcal{P} := \left\{\mathbf{p} = (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 \middle| p_1 + p_2 + p_3 + p_4 = 1, \sum_{i=1}^{4} \frac{(p_i - q_i)^2}{p_i} \leq 0.5, \mathbf{p} \geq 0\right\}.$$

Using this uncertainty set, we derive that the worst-case objective value for the nominal solution is 1.2; moreover, the non-adjustable robust counterpart of (4.29) is

$$\min_{\mathbf{d}} \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{4} \left[\left(1 + 5d_1 + 5d_2 + e_1^i - e_2^i\right)^2 + \left(1 + 5d_1 + 10d_2 + e_1^i + e_2^i\right)^2\right] p_i,$$

and its robust optimal solution is (-0.2, 0) with objective value 1.0. Next we shall examine three adjustable formulations; namely, linear, nonlinear, and cell-based decision rules.

Linear Decision Rule

Now we assume that the functions of the adjustable controllable factors $\mathbf{d}$ are linear in the observed values of the environmental factors $\mathbf{e}$. The *fully adjustable* linear decision rule is

$$d_j = D_j(\mathbf{x_j}, \mathrm{I}_2\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{j2}e_2 \quad \forall j \in \{1, 2\} \tag{4.31}$$

where $\mathrm{I}_2 = [1, 0; 0, 1]$. When both controllable factors are fully adjustable, all decisions are made only after the uncertainty is revealed. Consequently, the ARC of (4.29) with linear decision rule is

$$\min_{\mathbf{X}} \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{4} \Bigg[\Big(1 + 5D_1(\mathbf{x_1}, \mathbf{e^i}) + 5D_2(\mathbf{x_2}, \mathbf{e^i}) + e_1^i - e_2^i\Big)^2 + \Big(1 + 5D_1(\mathbf{x_1}, \mathbf{e^i}) + 10D_2(\mathbf{x_2}, \mathbf{e^i}) + e_1^i + e_2^i\Big)^2 \Bigg] p_i.$$

The optimal solution is $\mathbf{x_1}^*=(-0.2, -0.2, 0.6)^T$ and $\mathbf{x_2}^*=(0, 0, -0.4)^T$; i.e., the linear decision rules are $D_1(\mathbf{x_1}, \mathbf{e})=-(1 + e_1 - 3e_2)/5$ and $D_2(\mathbf{x_2}, \mathbf{e})=-2e_2/5$, so we have

$$\begin{aligned} 5d_1 + 5d_2 &= 5D_1(\mathbf{x_1}, \mathbf{e}) + 5D_2(\mathbf{x_2}, \mathbf{e}) = -1 - e_1 + e_2 \\ 5d_1 + 10d_2 &= 5D_1(\mathbf{x_1}, \mathbf{e}) + 10D_2(\mathbf{x_2}, \mathbf{e}) = -1 - e_1 - e_2. \end{aligned}$$

Therefore, the ARC yields the lowest possible optimal objective value; namely, zero for the problem, whereas the non-adjustable RC yields one.

More interesting cases have wait-and-see and here-and-now decisions together or at least one of the controllable factors is not fully adjustable. Table 4.2 presents the numerical results for all possible combinations of linear decision rules. The first column gives the possible linear decision rules for the two adjustable factors d_1 and d_2, where 'na' denotes *non-adjustable*, 'e1' denotes a factor that is adjustable on e_1; similarly, 'e2' denotes adjustability on e_2, and 'e[1,2]' denotes a *fully adjustable* factor. The second and third columns are the optimal coefficients (variables) $\mathbf{x_1}$ and $\mathbf{x_2}$ of the decision rules $D_1(.)$ and $D_2(.)$, where $(-)$ denotes a variable that vanishes in the associated decision rule. The final column (Obj.) presents the robust optimal objective value for the associated decision rule. Altogether, the numerical results show that when one of the factors is non-adjustable and the other is adjustable on

Table 4.2 – Linear Decision Rules (LDR) and Objective Values

LDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	(x_{10}, x_{11}, x_{12})	(x_{20}, x_{21}, x_{22})	
na	na	(-0.2,−, −)	(0,−, −)	1.00
na	e1	(-0.189,−, −)	(-0.009,-0.103,−)	0.66
na	e2	(-0.2,−, −)	(0,−,0)	1.00
e1	na	(-0.2,-0.2,−)	(0,−, −)	0.50
e2	na	(-0.2,−,0)	(0,−, −)	1.00
na	e[1,2]	(-0.186,−, −)	(-0.008,-0.12,-0.04)	0.62
e[1,2]	na	(-0.2,-0.2,0)	(0,−, −)	0.50
e1	e1	(-0.2,-0.2,−)	(0,−, −)	0.50
e1	e2	(-0.2,-0.2,0)	(0,−,-0.04)	0.45
e2	e1	(-0.21,−,0.04)	(-0.001,-0.09,−)	0.68
e2	e2	(-0.2,−,0.6)	(0,−,-0.4)	0.50
e[1,2]	e1	(-0.2,-0.2,0)	(0,−, −)	0.50
e[1,2]	e2	(-0.2,-0,2,0.6)	(−, −,0.4)	0.00
e1,	e[1,2]	(-0.2,-0.2,−)	(0,0,-0.04)	0.45
e2,	e[1,2]	(-0.2,−,0.6)	(0,-0.12,-0.4)	0.05
e[1,2]	e[1,2]	(-0.2,-0.2,0.6)	(0,0,-0.4)	0.00

e_2—see row (na, e2) or (e2, na) in Table 4.2—the optimal objective value of the ARC is the same as that of the non-adjustable RC. In all other cases the optimal objective value of the non-adjustable RC improves with at least 32% (see row (e2, e1)) for the ARC, and the highest improvement (100%) is attained when the first factor is fully adjustable and the second one is non-adjustable; see row (e[1,2], e2). Another interesting outcome is that introducing an adjustable factor into the problem may change the optimal decision for the non-adjustable factor; i.e., an optimal here-and-now factor can have different values in the ARC and RC. For example, if d_1 is adjustable on e_1 and d_2 is non-adjustable, then the optimal d_2 is -0.189 in the ARC, but it is -0.2 in the RC; see (na, na) and (na, e1).

Nonlinear Decision Rule

Table 4.3 shows the following nonlinear (quadratic) decision rule:

$$D_j(\mathbf{x_j}, \mathrm{I}_2\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{j2}e_2 + x_{11}^{(j)}e_1^2 + x_{22}^{(j)}e_2^2 + x_{12}^{(j)}e_1e_2 \tag{4.32}$$

where $\mathbf{x_j} = [x_{j0}, x_{j1}, x_{j2}, x_{11}^{(j)}, x_{12}^{(j)}, x_{22}^{(j)}]$. For example, if the controllable factors are only adjustable in e_1, then the decision rule is

$$D_j(\mathbf{x_j}, P_j\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{11}^{(j)}e_1^2$$

where $P_j = [1, 0; 0, 0]$. Obviously, the nonlinear decision rule is more general than the linear rule, so it is at least as good as the linear decision rule used in Table 4.2.

Table 4.3 – Nonlinear Decision Rules (NDR) and Objective Values

NDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	$(x_{10}, x_{11}, x_{12}, x_{11}^{(1)}, x_{22}^{(1)}, x_{12}^{(1)})$	$(x_{20}, x_{21}, x_{22}, x_{11}^{(2)}, x_{22}^{(2)}, x_{12}^{(2)})$	
na	na	(-0.2,-,-,-,-,-)	(0,-,-,-,-,-)	1.00
na	e1	(-0.196,-,-,-,-,-)	(0.014,0.093,-,-0.079,-,-)	0.65*
na	e2	(-0.2,-,-,-,-,-)	(-0.33,-,0,-,1.33,-)	1.00
e1	na	(-0.195,-0.2,-,-0.02,-,-)	(0,-,-,-,-,-)	0.50
e2	na	(-0.192,-,0,-,-,-)	(-,-0.03,-,-,-,-)	1.00
na	e[1,2]	(-0.188,-,-,-,-,-)	(0.062,-0.106,-0.053,-0.244,0.015,0.05)	0.58*
e[1,2]	na	(-0.178,-0.2,0,-0.044,-0.044,0)	(0,-,-,-,-,-)	0.50
e1	e1	(-0.213,-0.2,-,0.05,-,-)	(-0.002,0,-,0.008,-,-)	0.50
e1	e2	(-0.533,-0.2,-,1.33,-,-)	(-0.005,-,-0.04,-,0.02,-)	0.45
e2	e1	(-0.19,-,0.044,-,-0.047,-)	(-0.045,-0.063,-,0.197,-,-)	0.65*
e2	e2	(-0.223,-,0.6,-,0.094,-)	(-0.002,-,-0.4,-0.008,-)	0.50
e[1,2]	e1	(-0.186,-0.2,0,-0.028,0.028,0)	(-0.006,-,-,0.023,-,-)	0.50
e[1,2]	e2	(-0.253,-0.2,0.6,0.107,0.107,0)	(-0.002,-,-0.4,-,0.008,-)	0.00
e1	e[1,2]	(-0.146,-0.199,-,-0.214,-,-)	(-0.052,-0.003,-0.047,0.148,0.057,-0.001)	0.44*
e2	e[1,2]	(-0.233,0,0.6,-,0.133,0)	(-0.151,-0.12,-0.4,0.105,0.5,-)	0.05
e[1,2]	e[1,2]	(-0.214,-0.2,0.6,0.028,0.028,0)	(0.013,0,-0.4,-0.066,0.014,0)	0.00

(∗) denotes an improved optimal objective value compared with that in Table 4.2

We denote the cases where the nonlinear decision rule performs better than the linear by (*) in the last column of Table 4.3. The highest improvement compared with Table 4.2 is obtained when the first factor is non-adjustable and the second factor is fully adjustable; compare (na, e[1,2]) in Table 4.3, where the optimal objective value is 0.584, with the objective 0.621 for the same situation in Table 4.2. Quantifying the *value of information* is an important topic in both adjustable robust decision making and in general decision making; Table 4.3 shows that having extra information on e_2 but not on e_1 for one of the controllable factors has no added value in the adjustable decision making; i.e., the non-adjustable and the adjustable RCs have the same

optimal objective at (na, e2) and (e2, na). On the other hand, having information on e_1 for one of the controllable factors yields improvement in the objective; see (na, e1) and (e1, na) in Table 4.3. Moreover, if d_1 responds to both environmental factors, and d_2 uses information on e_2 only, then we obtain the lowest possible optimal objective value (namely, zero); see (e[1,2], e2) in Table 4.3.

Cell-Based Decision Rule

Now we propose a new type of decision rule that we call the *cell-based* decision rule:

$$D_j(\mathbf{x_j}, \mathbf{e}) := x_{ji} \text{ if } \mathbf{e} \in \text{cell}(i), i \in V_j, \tag{4.33}$$

where $\mathbf{x_j}$ is the decision vector for the jth adjustable variable (x_{ji} being the decision for the ith cell), cell(i) is the region determined by the ith cell, and V_j is the set of cell indices for the associated information-base.

Remark 11 *Cells used in the decision rule are non-intersecting squares in two dimensions, cubes in three dimensions, and hypercubes in higher dimensional uncertainty spaces. For the sake of simplicity, we assume that all cells have the same volume (unless some cells are not merged to include enough data points).*

If in the illustrative example (4.29), d_1 is fully adjustable and the uncertainty set is divided into four cells, then the decision rule is

$$D_1(\mathbf{x_1}, \mathbf{e}) := \begin{cases} x_{11}, & \mathbf{e} \in \text{cell}(1) \\ x_{12}, & \mathbf{e} \in \text{cell}(2) \\ x_{13}, & \mathbf{e} \in \text{cell}(3) \\ x_{14}, & \mathbf{e} \in \text{cell}(4) \end{cases} \tag{4.34}$$

where $\text{cell}(i) := \{(e_1, e_2) \in \mathbb{R}^2 : \ell_{1i} \leq e_1 \leq u_{1i}, \ell_{2i} \leq e_2 \leq u_{2i}\}$ ($i \in \{1, 2, 3, 4\}$), $\boldsymbol{\ell_1} = [0, -1, -1, 0]$, $\boldsymbol{u_1} = [1, 0, 0, 1]$, $\boldsymbol{\ell_2} = [0, 0, -1, -1]$, $\boldsymbol{u_2} = [1, 1, 0, 0]$, and $V_1 = \{1, 2, 3, 4\}$; cells are represented by their center points in (4.30). To show the difference between full and partial information, we assume that d_1 is adjustable on e_1 but not on e_2. The associated decision then becomes

$$D_1(\mathbf{x_1}, e_1) := \begin{cases} x_{11}, & e_1 \in \text{cell}(1) \\ x_{12}, & e_1 \in \text{cell}(2) \end{cases} \tag{4.35}$$

where $\text{cell}(1) := \{e_1 \in \mathbb{R} : 0 \leq e_1 \leq 1\}$ and $\text{cell}(2) := \{e_1 \in \mathbb{R} : -1 \leq e_1 \leq 0\}$, and $V_1 = \{1, 2\}$). It is easy to see that (4.35) implies that when e_2 is extracted from the information-base, the new cells are projections from the cells in the two-dimensional space in (4.34) onto the one-dimensional space on e_1 in (4.35). The disadvantage of the cell-based decision rule is that this rule often has more variables compared with

the linear and nonlinear decision rules, especially when the number of cells is high. Nevertheless the numerical results for the example show that the new decision rule is better than the linear, and is 'almost' as good as the nonlinear decision rule—even when the total number of cells is only four; see Table 4.4.

Table 4.4 – Cell-Based Decision Rules (CDR) and Objective Values

CDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	$(x_{11}, x_{12}, x_{13}, x_{14})$	$(x_{21}, x_{22}, x_{23}, x_{24})$	
na	na	(-0.2,−, −, −)	(0,−, −, −)	1.00
na	e1	(-0.19,−, −, −)	(-0.06,0.04,−, −)	0.66
na	e2	(-0.2,−, −, −)	(0,0,−, −)	1.00
e1	na	(-0.3,-0.1,−, −)	(0,−, −, −)	0.50
e2	na	(-0.2,-0.2,−, −)	(0,−, −, −)	1.00
na	e[1,2]	(-0.19,−, −, −)	(-0.09,0.03,0.07,-0.05)	0.62
e[1,2]	na	(-0.3,-0.1,-0.1,-0.3)	(0,−, −, −)	0.50
e1	e1	(-0.3,-0.1,−, −)	(0,0,−, −)	0.50
e1	e2	(-0.3,-0.1,−, −)	(-0.02,0.02,−, −)	0.45
e2	e1	(-0.18,-0.2,−, −)	(-0.06,0.04,−, −)	0.65
e2	e2	(0.1,-0.5,−, −)	(-0.2,0.2,−, −)	0.50
e[1,2]	e1	(-0.3,-0.1,-0.1,-0.3)	(0,0,−, −)	0.50
e[1,2]	e2	(0,0.2,-0.4,-0.6)	(-0.2,0.2,−, −)	0.00
e1	e[1,2]	(-0.3,-0.1,−, −)	(-0.02,-0.02,0.02,0.02)	0.45
e2	e[1,2]	(0.1,-0.5,−, −)	(-0.26,-0.14,0.26,0.14)	0.05
e[1,2]	e[1,2]	(0,0.2,-0.4,-0.6)	(-0.2,-0.2,0.2,0.2)	0.00

To the best of our knowledge, parametric decision rules in the RO literature cannot handle adjustable *integer* variables, since the adjustable decision is a function of the uncertain parameter $\mathbf{e}$, and the function does not necessarily take integer values for all $\mathbf{e}$; see (4.31) and (4.32). However, the cell-based decision rule can handle such variables. As we can see from (4.33), the adjustable decision x_{ji} can take integer values since the cell-based decision rule relates $\mathbf{e}$ and x_{ij} through an 'if' statement. Therefore, if we make x_{ij} an integer variable, then the cell-based decision rule gives integer decisions. Note that similar decision rules have also been proposed by Bertsimas and Caramanis (2010); Hadjiyiannis et al. (2011); Vayanos et al. (2011).

Now using the illustrative example, we show the validity of our approach for adjustable integer variables. We modify the old example in the following way: we

assume d_1 and d_2 are adjustable, and they take values that are multiples of 1/4.

Table 4.5 – Integer Cell-Based Decision Rules and Objective Values

CDR		$D_1(.)$	$D_2(.)$	Obj.
d_1	d_2	$(x_{11}, x_{12}, x_{13}, x_{14})$	$(x_{21}, x_{22}, x_{23}, x_{24})$	
na	na	(-0.25,−, −, −)	(0,−, −, −)	1.41
na	e1	(-0.25,−, −, −)	(0,0,−, −)	1.41
na	e2	(-0.25,−, −, −)	(0,0,−, −)	1.41
e1	na	(-0.25,0,−, −)	(0,−, −, −)	0.92
e2	na	(-0.25,-0.25,−, −)	(0,−, −, −)	1.41
na	e[1,2]	(-0.25,−, −, −)	(0,0,0,0)	1.41
e[1,2]	na	(-0.25,0,0,-0.25)	(0,−, −, −)	0.92
e1	e1	(-0.25,0,−, −)	(0,0,−, −)	0.92
e1	e2	(-0.25,0,−, −)	(0,0,−, −)	0.92
e2	e1	(-0.25,-0.25,−, −)	(0,0,−, −)	1.41
e2	e2	(0.25,-0.5,−, −)	(-0.25,0.25,−, −)	1.29
e[1,2]	e1	(-0.25,0,0,-0.25)	(0,0,−, −)	0.92
e[1,2]	e2	(0,0.25,-0.50,-0.75)	(-0.25,0.25,−, −)	0.26
e1	e[1,2]	(-0.25,0,−, −)	(0,0,0,0)	0.92
e2	e[1,2]	(0,-0.50,−, −)	(-0.25,0,0.25,0.25)	0.99
e[1,2]	e[1,2]	(0,0.25,-0.50,-0.75)	(-0.25,-0.25,0.25,0.25)	0.26

Table 4.4 presents optimal decision rules and resulting objective values for all possible combinations of adjustability. These numerical results show that we obtain important improvements for the non-adjustable (na, na) optimal objective value by using the cell-based decision rule. As may be anticipated, the integer formulation yields higher (worse) optimal objective values. For example, we can no longer get a zero objective in Table 4.5. Moreover, in contrast to the continuous case, we can no longer improve the optimal objective of the non-adjustable RC at decision rule combinations: (na, e1), (na, e[1,2]), and (e2, e1); see the corresponding rows of Tables 4.4 and 4.5.

4.5 Realistic Examples

In this section, we present realistic examples to demonstrate the effectiveness of our methods. We use a 64-bit Windows PC with a 2.2 GHz Intel Core i7 processor, and 8 GB of RAM. To solve the mathematical optimization problems, we use KNITRO 8.0 embedded in MATLAB (2012a) and AIMMS 3.12.

4.5.1 Television Images

In color televisions the quality of signals is determined by the power signal-to-noise-ratios in the image transmitted. We take the response function $y(\mathbf{d}, \mathbf{e})$ from Myers et al. (2009, p. 512), where the response y measures the quality of transmitted signals in decibels. The controllable factors are the number of tabs in a filter d_1, and the sampling frequency d_2; the environmental factors are the number of bits in an image e_1, and the voltage applied e_2. The least-square estimate of the metamodel is

$$\begin{aligned}\hat{y}(\mathbf{d}, \mathbf{e}) \ &= 33.389 - 4.175d_1 + 3.748d_2 + 3.348d_1d_2 - 2.328d_1^2 - 1.867d_2^2 \\ &\quad - 4.076e_1 + 2.985e_2 - 2.324d_1e_1 + 1.932d_1e_2 + 3.268d_2e_1 - 2.073d_2e_2\end{aligned}$$

where all factors are coded; for details on the DoE we refer to Myers et al. (2009, pp. 511–515).
We find the optimal design settings of d_1 and d_2 using the optimization problem (4.9):

$$\begin{aligned}&\max_{\mathbf{d}} \ \mathrm{E}_{\mathbf{e}}[\hat{y}(\mathbf{d}, \mathbf{e})] \\ &\text{s.t. } \mathrm{Var}_{\mathbf{e}}[\hat{y}(\mathbf{d}, \mathbf{e})] \leq T.\end{aligned} \tag{4.36}$$

The robust counterpart of (4.36) is given by Theorem 1 in §4.3.2. To estimate $\mathbf{q}$, we use the historical data in Figure 1. Since we have no real data, we have randomly created these data. The sample size is $N = 350$, and the support of $\mathbf{e}$ is divided into 25 cells of the same volume so $V = \{1, \ldots, 25\}$, $\mathbf{q} = [q_1, \ldots, q_{25}]^T$ and $\rho = \chi^2_{0.999,24}/350$; see (4.11). We shall use the same data in our two realistic examples so the data do not favor our method.

The goal of these examples is to compare the optimal solutions of the nominal and robust counterpart problems. In §4.5.1.1 we shall compare the worst-case and average performances of these two solutions via the objective value and the constraint violation. In §4.5.1.2 we shall compare the confidence levels probabilities of the nominal and robust optimal solutions.

4.5.1.1 Robust versus Nominal Solutions

We vary the right-hand side value T of the variance constraint from 0.1 to 0.8; see column one in Table 4.6. We solve the nominal and the RC problems for these T values, and compare the worst-case performances of the nominal and robust optimal solutions.
Columns two and three are the robust optimal solution (d_1^*, d_2^*) and its objective value (y^*). Column four (Var.) is the robust variance of the response. Column five

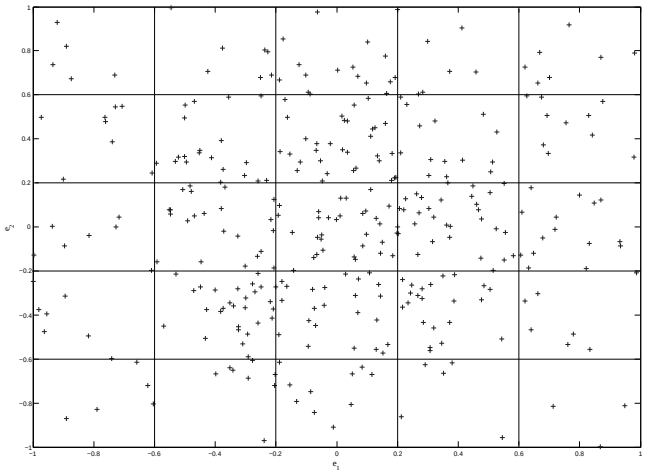


Figure 4.1 – Historical Data on **e**

Table 4.6 – Worst-Case Analysis for TV Image Example

T	Robust			Nominal		
	(d_1^*, d_2^*)	y^*	Var.(%)	(d_1^*, d_2^*)	W-C($\mathbb{E}$)	W-C(Var)(%)
0.1	(-0.43,0.83)	35.27	0.10(0%)	(-0.4472,0.7755)	35.27	0.19(88%)
0.2	(-0.44,0.79)	35.28	0.15(-25%)	(-0.4625,0.6853)	35.26	0.37(87%)
0.3	(-0.44,0.79)	35.28	0.15(-50%)	(-0.4763,0.6152)	35.23	0.56(86%)
0.4	(-0.44,0.79)	35.28	0.15(-62%)	(-0.4867,0.5648)	35.20	0.71(79%)
0.5	(-0.44,0.79)	35.28	0.15(-70%)	(-0.4867,0.5648)	35.20	0.71(43%)
0.6	(-0.44,0.79)	35.28	0.15(-75%)	(-0.4867,0.5648)	35.20	0.71(19%)
0.7	(-0.44,0.79)	35.28	0.15(-78%)	(-0.4867,0.5648)	35.20	0.71(2%)
0.8	(-0.44,0.79)	35.28	0.15(-81%)	(-0.4867,0.5648)	35.20	0.71(-11%)

is the nominal optimum solution. Column six presents the mean ($\mathbb{E}_{\mathbf{e}}[y(\mathbf{e}, \mathbf{d})]$) when the decision factors are fixed to the nominal optimum solution $\mathbf{d}^*$ and the worst-case probability vector $\mathbf{p}^*$ that minimizes the expectation is realized. Column seven (W.-C.(Var)) gives the worst-case value of the variance for the nominal solution; now we consider the probability vector $\mathbf{p}^*$ that maximizes the variance, as the worst-case. Notice that we also report the percentages of the worst-case constraint violations for the nominal and robust optimal solutions; these values are shown within brackets (%) in the columns four and seven. We use the formula [(W.-C.(Var)-T) $\times$ $100/T$] to calculate the constraint violation percentage of the nominal solution when the worst-case uncertain parameters are realized. If the violation percentage is negative, then the constraint is satisfied; 0% means the constraint is binding (see column four,

row T=0.1); a positive percentage means constraint violation at column seven.

Table 4.6 reveals that the robust reformulation of the variance constraint becomes redundant when $T \geq 0.2$; e.g., the left-hand side value of the constraint for the robust optimal solution (-0.4439,0.7988) when $T = 0.2$ is 0.149. Therefore we have the same robust optimal solution for $T \geq 0.2$. The nominal variance constraint (not the worst-case) also becomes redundant when $T \geq 0.389$, and therefore we have the same nominal optimum solution when $T \geq 0.4$. Notice that the robust optimal objective values and the worst-case objective values of the nominal optimum solutions are within 1% of each other. However, the constraint violation percentages favor our robust approach; e.g., the percentages in column seven show that when $T \leq 0.4$, the nominal optimal solution violates the constraint on average 85% in the worst-case. When $T \geq 0.715$ the nominal optimum solution no longer violates the constraint in the worst-case, but it is closer to be binding than the robust solution. All together, using our robust optimization method for this example, we gain immunity to the worst-case uncertainty without being penalized by the objective.

Now we analyze the average mean and variance of the response for the nominal and robust optimal solutions, by randomly sampling 1,000 probability vectors $\mathbf{p}$ from the uncertainty set that is used in the RC; see (4.12). We sample as follows: First we uniformly sample 25 individual probabilities ($\hat{p}_i$). Then to guarantee that the probabilities sum up to one, we use the adjustment: $p_i = \hat{p}_i / \sum_{j=1}^{25} \hat{p}_j$. Finally, if the probability vector $\mathbf{p}$ is within the uncertainty set, we keep $\mathbf{p}$ in the sample; else we discard $\mathbf{p}$ and sample again. We repeat this procedure until we have 1,000 accepted vectors that are within the set. The confidence level of the uncertainty set (4.12) is 99.9%. From the $\mathbf{p}$ in the sample we estimate the mean in (4.14) and the variance in (4.16) for a given solution, and then take the averages; see Table 4.7.

Table 4.7 – Average (Avg) Mean ($\mathbb{E}$) and Variance (Var)

T	Robust		Nominal	
	Avg($\mathbb{E}$)	Avg(Var)	Avg($\mathbb{E}$)	Avg(Var)
0.1	35.37	0.064	35.41	0.12
0.2	35.40	0.095	35.45	0.24
0.3	35.40	0.095	35.46	0.35
0.4	35.40	0.095	35.47	0.46

Table 4.7 shows that the average expected response for the nominal solutions are

less than 1% better than that of the robust solutions (we maximize $\mathbb{E}(y)$; see (4.29)). Nevertheless, the average response variances of the robust optimal solutions are more than 100% better (smaller) than that of the nominal solutions when $T \leq 0.2$, and the improvement becomes more than 350% when $T \geq 0.3$. Moreover, we found that on average the nominal solutions violate 5% of the variance constraint, whereas the robust solutions are not at all binding. The effectiveness of our approach is further analyzed in the following subsection.

4.5.1.2 Confidence Level

The uncertainty set (4.12) coincides with the $(1-\alpha)$-confidence set for the unknown probability vector $\mathbf{p}$ centered at the empirical estimate $\mathbf{q}$. In the numerical examples we use $\alpha = 0.001$, which means that the robust optimal solution is immune to uncertainty with at least 99.9% probability. These uncertainties are ignored in the nominal problem; nevertheless, the nominal solution may be robust for some part of the uncertainty set. To find the confidence level of a nominal solution, we calculate the tightest uncertainty set for which the nominal solution at hand is robust; i.e., we adjust the ρ value in (4.12); see Table 4.8.

Table 4.8 – Confidence Levels (1-α) of Nominal Solutions

T	≤ 0.4	0.5	0.6	0.7	≥ 0.8
$(1-\alpha)$	0%	2%	70%	98%	99.9%

Table 4.8 shows that when $T \leq 0.4$ the nominal solutions have no immunity to uncertainty at all; i.e., a 'small' change in the given empirical estimate $\mathbf{q}$ results in infeasibility in the nominal variance constraint. We may anticipate this result, since the nominal solutions are binding for the associated cases. On the other hand, for $T = 0.5$ to $T = 0.7$ we see important improvement in the immunity to uncertainty for the nominal solution. In addition, when $T \geq 0.8$, the immunity of the nominal optimum solution is as good as the robust optimum solution—even though the nominal optimum solution is closer to be binding in the constraint than the robust solution. Concerning the objective, we have already shown in the last row of Table 4.6 that the robust and the 'worst-case' nominal optimum objective values are almost the same; i.e., the robust solutions are less than 1% better than the nominal solutions. In conclusion, the robust reformulation improves the immunity to uncertainty when $T \leq 0.8$ and the improvement is even better when $T \leq 0.4$. Additionally, we prefer the robust solution when $T \geq 0.8$, since it performs better in the worst-case.

4.5.2 Distribution Center

In this example we focus on the simulation of a cross-docking distribution center (CDDC) developed by Shi (2011). The CDDC is used by a Chinese automobile manufacturer that needs to improve the physical flow in its supply chain. The objective of the company is steady production. The challenge in such production is the supply uncertainty due to environmental factors, mentioned below. To model the associated supply chain system, Shi proposes a hybrid approach that combines discrete-event simulation, RSM, and Taguchi's world view. We focus on Shi's linear regression metamodel for the total throughput with five controllable factors; namely, the number of receiving doors d_1, the number of shipping doors d_2, the number of forklifts d_3, the number of conveyors d_4, and the supply chain threshold in storage d_5. The model has two environmental factors; namely, the variability in less-than-truckload supply shipments e_1, and the production interruption or delay probabilities of the suppliers e_2. Using the simulation's I/O data, Shi approximates the unknown I/O function of the total throughput by the following least-squares estimate $\hat{y}$:

$$\begin{aligned}\hat{y}_{\mathrm{TT}}(\mathbf{e},\mathbf{d}) =& 10^4 \times \Big[47.97 + 3.982d_1 + 2.025d_2 - 0.031d_3 + 0.734d_4 + 0.034d_5 \\ &+ 0.789d_1d_2 + 0.012d_1d_3 + 0.003d_1d_4 - 0.002d_1d_5 + 0.0007d_2d_3 \\ &- 0.065d_2d_4 - 0.1131d_2d_5 - 0.078d_3d_4 - 0.041d_3d_5 + 0.11d_4d_5 \\ &- 3.406d_1^2 - 1.781d_2^2 + 0.011d_3^2 - 1.033d_4^2 + 0.111d_5^2 \\ &+ (16.66 + 1.511d_1 + 2.374d_2 - 0.059d_3 + 0.824d_4 - 0.093d_5)e_1 \\ &- (0.005 + 0.27d_1 + 0.661d_2 - 0.086d_3 + 0.335d_4 - 0.005d_5)e_2\Big],\end{aligned}$$

where all factors are coded such that $-1 \leq \mathbf{d} \leq 1$, and $-1 \leq \mathbf{e} \leq 1$. More precisely, the coded controllable factors are between -1 and 1 because of the physical restrictions of the production facility. Shi's ANOVA shows that the metamodel $\hat{y}_{\mathrm{TT}}$ have non-significant lack-of-fit; and for the estimated parameters the level-of-significance is 0.05. Using Shi's response model, we focus on the robust reformulation of the following optimization problem:

$$\begin{aligned}\min_{\mathbf{1}\leq\mathbf{d}\leq\mathbf{1}} \quad & \mathrm{Var}_{\mathbf{e}}[\hat{y}(\mathbf{d},\mathbf{e})] \\ \text{s.t.} \quad & E_{\mathbf{e}}[\hat{y}[\mathbf{d},\mathbf{e})] \geq T.\end{aligned} \tag{4.37}$$

To estimate the frequencies $\mathbf{q}$ used by the nominal and RC problems, we use the same historical data as in Figure 4.1.

Table 4.9 compares the worst-case performances of the nominal and robust optimal solutions. Besides the worst-case mean and standard deviation, we construct the

Table 4.9 – Worst-Case Comparison for CDDC Example

T	Robust			Nominal (Worst-Case)		
	$\sigma_{\mathbf{e}}(\hat{y})$	$\mathbb{E}_{\mathbf{e}}[\hat{y}]$	$\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\hat{\sigma}_{\mathbf{e}}$	$\sigma_{\mathbf{e}}(\hat{y})$	$\mathbb{E}_{\mathbf{e}}[\hat{y}]$	$\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\hat{\sigma}_{\mathbf{e}}$
$\leq$339	67	339	(136, 541)	67	339	(136, 541)
340	67	340	(137, 542)	67	339	(136, 541)
370	70	370	(158, 581)	68	348	(143, 552)
400	74	400	(176, 623)	71	377	(163, 590)
420	79	420	(182, 657)	73	396	(174, 617)
430	81	430	(184, 675)	75	405	(178, 632)
440	85	440	(184, 695)	77	414	(181, 648)
450	88	450	(183, 716)	80	424	(183, 664)

$***$ all entries should be multiplied by 1,000

worst-case confidence interval for $\hat{y}$; namely, $\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\hat{\sigma}_{\mathbf{e}}$. Because the probability distribution of the environmental factors is unknown, we cannot say much about the confidence level of this interval. The numerical results show that in the worst-case the nominal solutions are on average 6% lower than the target T for the expected total throughput; i.e., the average violation of the constraint is 6%. The worst-case standard deviations of the total throughput for the nominal solutions are on average 8% lower than that of the robust ones. Nevertheless the confidence intervals for the robust optimal solutions are always shifted to the right compared with the nominal solutions, which is in favor of the robust approach since a higher total throughput is better. Altogether the numerical results favor the robust approach when $T >$339000; when $T \leq$ 339000 both methods yield the same outcome.

In Table 4.10 we compare the average performance of robust and nominal solutions via Monte Carlo simulation. First, using the given historical data in Figure 4.1 as the nominal value ($\mathbf{q}$) of the uncertainty set, we generate 1,000 probability vectors, like we did for the TV images example. Then we calculate the expected response and variance for the nominal and robust solutions at each probability vector, and take the averages. Table 4.10 shows higher means and lower variances than Table 4.9 in the worst-case. Table 4.9 is based on the worst-case scenario; i.e., we maximize the variance and minimize the expectation; Table 4.10 is based on the average performance. Table 4.10 shows that the confidence intervals for the robust solutions are always shifted to the right compared with the nominal solutions, so we prefer the robust approach, however, the robust solutions have slightly higher variations (from

Table 4.10 – Average Comparison for CDDC Example (Simulation Results)

T	Robust			Nominal		
	$\bar{\sigma}_{\mathbf{e}}(\hat{y})$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}] \pm 3\bar{\sigma}_{\mathbf{e}}$	$\bar{\sigma}_{\mathbf{e}}(\hat{y})$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}] \pm 3\bar{\sigma}_{\mathbf{e}}$
$\leq$339	58	361	(186, 536)	58	360	(185, 535)
340	58	361	(186, 536)	58	360	(185, 535)
370	61	392	(209, 575)	59	369	(192, 547)
400	64	423	(229, 618)	61	399	(214, 585)
420	68	445	(238, 652)	64	419	(227, 612)
430	71	456	(241, 670)	65	429	(232, 627)
440	74	467	(244, 690)	67	439	(236, 643)
450	77	478	(245, 711)	69	449	(239, 659)

$***$ all entries should be multiplied by 1,000

0% to 10%) than the nominal solutions.

Adjustable Robust Optimization

One of the most important decision factors in the CDDC is the number of shipping doors d_2. A moderate number of shipping doors may increase the inventory in the temporary storage area, whereas an excessive number causes a low utilization of doors. We now assume that d_2 is adjustable according to the uncertain parameter e_1—namely, the variability in less-than-truckload supply shipments. More precisely, we assume that the number of shipping doors can be adjusted after the variability in supply shipments has been observed. We use the cell-based decision rule introduced in §4.4; the data and the 25 cells are presented in Figure 4.1. Notice that the domain of e_1 is divided into five equal intervals; for each interval we have a different decision. In Table 4.11, columns two through four present the actual 'worst-case' performance of the ARC, and columns five through eight present the average performance of adjustable robust solutions via Monte Carlo simulation as in Table 4.10. We use the same probability vector sample as in the television images example. These numerical results show that the worst-case confidence intervals of the ARC are tighter than those of the general RC in Table 4.9. This is because of the improved response variances of the adjustable robust solutions, the improvement percentages (%) are reported in column two within brackets; e.g., it is as high as 10% when T =450000. As we anticipated, the simulation results show that the average performance of the adjustable robust solutions is better than the worst-case; i.e., the average of the response mean $\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$ is higher than the worst-case mean $\mathbb{E}_{\mathbf{e}}[\hat{y}]$, and the average variance $\bar{\sigma}_{\mathbf{e}}(\hat{y})^2$ is lower than the worst-case variance $\sigma_{\mathbf{e}}(\hat{y})^2$. Comparing the average perfor-

Table 4.11 – ARO Results for CDDC Example

	Worst-Case			Average		
T	$\sigma_{\mathbf{e}}(\hat{y})$	$\mathbb{E}_{\mathbf{e}}[\hat{y}]$	$\mathbb{E}_{\mathbf{e}}[\hat{y}] \pm 3\sigma_{\mathbf{e}}$	$\bar{\sigma}_{\mathbf{e}}(\hat{y})$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}]$	$\bar{\mathbb{E}}_{\mathbf{e}}[\hat{y}] \pm 3\bar{\sigma}_{\mathbf{e}}$
$\leq$339	67 (0%)	339	(136, 541)	58	367	(192, 541)
340	67 (0%)	340	(137, 542)	58	367	(193, 542)
370	68 (2%)	370	(167, 572)	59	389	(194, 543)
400	71 (3%)	400	(188, 611)	61	418	(212, 566)
420	73 (6%)	420	(201, 639)	63	425	(235, 615)
430	74 (7%)	430	(209, 650)	64	448	(246, 630)
440	75 (10%)	440	(215, 665)	65	449	(254, 644)
450	78 (10%)	450	(216, 684)	69	456	(249, 661)

$***$ all entries should be multiplied by 1,000

mances of the non-adjustable and adjustable robust solutions in Tables 4.10 and 4.11 shows that the ARC yields tighter confidence intervals that are subintervals of the confidence intervals in the general RC. Therefore, the ARC reduces the response uncertainty compared with the general RC. Finally, additional experimentation showed that making 'only' d_3, d_4 or d_5 adjustable on e_1 has an improvement less than 1% in the objective; i.e., in those cases the non-adjustable solutions are as good as the adjustable solutions. This shows that the quantitative value of information may significantly change for different parameters in ARO.

4.6 Conclusions and Future Research

In this chapter, we proposed a RO methodology for RPD, this methodology uses an uncertainty set based on historical data on the environmental variables; data may be collected from either real or simulated systems. Adding RO to RPD has the following advantages: (i) Unlike the classical RPD, we do not make any distributional assumptions on the uncertain parameters. (ii) We do account for the ambiguity caused by the lack of knowledge about distributions by using the so-called ϕ-divergence uncertainty sets based on historical data. (iii) Both RO and ARO methods are computationally tractable; ARO is tractable even for nonlinear decision rules. (iv) Our ARO approach can be used for modeling adjustable integer decision variables. In future research, we shall apply the cell-based decision rules to general classes of optimization problems with moderate numbers of 'integer' variables.

Bibliography

Al-Aomar, R. 2006. Incorporating robustness into genetic algorithm search of stochastic simulation outputs. *Simulation Modelling Practice and Theory* **14**(3) 201–223.

Angün, E. 2011. A risk-averse approach to simulation optimization with multiple responses. *Simulation Model. Practices and Theory* **19**(3) 911–923.

Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Math. Oper. Res.* **23**(4) 769–805.

Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Oper. Res. Lett.* **25**(1) 1–13.

Ben-Tal, A., D. den Hertog, J.-P. Vial. 2014. Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Prog.* forthcoming.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization.* Princeton Press, Princeton, NJ.

Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* **59**(2) 341–357.

Bertsimas, D., C. Caramanis. 2010. Finite Adaptability in Multistage Linear Optimization. *IEEE Transactions on Automatic Control* **55**(12) 2751–2766.

Bertsimas, D., O. Nohadani, K. M. Theo. 2010. Robust optimization for unconstrained simulation-based problems. *Oper. Res.* **58**(1) 161–178.

Bingham, D., V. N. Nair. 2012. Noise variable settings in robust design experiments. *Technometrics* **54**(4) 388–397.

Chan, Y.-J., D. J. Ewins. 2010. Management of the variability of vibration response levels in mistuned bladed discs using robust design concepts. *Mech. Systems and Signal Processing Part 1: Parameter Design* **24**(8) 2777–2791.

Dasgupta, T., C. F. J. Wu. 2006. Robust Parameter design with feed-back control. *Technometrics* **48**(3) 349–359.

Dellino, G., J. P. C. Kleijnen, C. Meloni. 2012. Robust optimization in simulation: Taguchi and Krige combined *INFORMS J. Comput.* **24**(3) 471–484.

Delpiano, J., M. Sepulveda. 2006. Combining iterative heuristic optimization and

uncertainty analysis methods for robust parameter design. *Engineering Optim.* **38**(7) 821–831.

El-Ghaoui, L., H. Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM. J. Matrix Anal. & Appl.* **18**(4) 1035–1064.

Fu, M. C. 2007. Are we there yet? The marriage between simulation & optimization. *OR/MS Today* **34**(3) 16–17.

Fu, M., B. Nelson. 2003. Simulation optimization. *ACM Trans. Model. and Simulation* **13**(2) 105–107.

Hadjiyiannis, M. J., P. Goulart, D. Kuhn. 2011. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization *Proceedings of the Decision and Control and European Control, 50th IEEE Conference* 7386–7391.

Hu, Zhaolin, J. Cao, L. J. Hong. 2012. Robust simulation of global warming policies using the DICE model. *Management Sci.* **58**(13) 2190–2206.

Iman, R. L., J. C. Helton. 2006. An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal.* **8**(1) 71–90.

Joseph, V. R. 2003. Robust parameter design with feed-forward control. *Technometrics* **45**(4) 284–292.

Kleijnen, J. P. C. 2008. *Design and Analysis of Simulation Experiments*. Springer-Verlag, Heidelberg, Germany.

Miranda, A. K., E. Del Castillo. 2011. Robust parameter design optimization of simulation experiments using stochastic perturbation methods. *J. Oper. Res. Soc.* **62**(1) 198–205.

Myers, R. H., D. C. Montgomery, C. M. Anderson-Cook. 2009. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd ed. John Wiley & Sons, Hoboken, NJ.

Nair, V. N., B. Abraham, J. MacKay, G. Box, R. N. Kacker, T. J. Lorenzen, J. M. Lucas. 1992. Taguchi's parameter design: a panel discussion. *Technometrics* **34**(2) 127–161.

Pardo, L. 2006. *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC, Boca Raton, FL.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.

Sanchez, S. M. 2000. Robust design: seeking the best of all possible worlds. *Proceedings of the 2000 Winter Simulation Conference* (edited by J.A. Joines, R.R. Barton, K Kang, and P.A. Fishwick) **1** 69–76.

Shi, W. 2011. Design of pre-enhanced cross-docking distribution center under supply uncertainty: RSM robust optimization method. Working Paper, Huazhong

University of Science & Technology, China.

Spall, J. 2003. *Introduction to Stochastic Search and Optimization.* Wiley-Interscience, New York, NY.

Stinstra, E., D. den Hertog. 2008. Robust optimization using computer experiments. *EJOR* **191**(3) 816–837.

Taguchi, G. 1986. *Introduction to Quality Engineering.* Asian Productivity Organization, UNIPUB, White Plains, NY.

Taguchi, G. 1987. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost.* UNIPUB/Kraus International, White Plains, NY.

Urval, R., S. Lee, S. V. Atre, S.-J. Park, R. M. German. 2010. Optimisation of process conditions in powder injection moulding of microsystem components using robust design method. *Powder Metallurgy Part 2: Secondary Design Parameters* **53**(1) 71–81.

Vanlı, O. A., E. Del Castillo. 2009. Bayesian approaches for on-line robust parameter design. *IIE Transactions* **41**(4) 359–371.

Vayanos, P., D. Kuhn, B. Rustem. 2011. Decision rules for information discovery in mutli-stage stochastic programming. *Proc. 50th IEEE Conf. on Decision and Control and European Control Conf. (CDC-ECC)*, Orlando, FL, 7368–7373.

Wiedemann, M. 2010. Robust Parameter Design for Agent-Based Simulation Models With Application in a Cultural Geography Model. Master's Thesis, Naval Postgraduate School, Monterey, CA.

Yanıkoğlu, İ., D. den Hertog. 2013. Safe approximations of ambiguous chance constraints using historical data. *INFORMS J. Comput.* **25**(4) 666–681.

CHAPTER 5

Robust Pessimistic Bilevel Optimization

5.1 Introduction

> *"Good optimization gives you the best choice based on the data. Great optimization is robust and resilient in the face of change and data errors."*
> *Sashihara (2011)*

Optimization problems often contain uncertain parameters because of measurement, estimation, and implementation errors. There are two novel optimization approaches for handling such errors, namely, *robust* and *stochastic* optimization. To model uncertain parameters, *robust optimization* (RO) uses uncertainty sets; and *stochastic optimization* (SO) uses probability distributions. Due to this basic difference, the two methods may significantly differentiate from each other. In this chapter we work on a class of uncertain optimization problems, where SO and RO coincide with each other, i.e., the associated modelling issue is not important when certain assumptions on uncertainty hold. However, in the sequel of the chapter, we generally adopt the RO terminology, and we shall use some of the techniques that are mainly invented in the realm of RO, and later extended to SO.

RO is a relatively young and active research field that is mainly developed in the course of the last 15 years. As it is explained above, the goal of RO is to find solutions that are robust with respect to the uncertainty of the problem parameters in a given mathematical optimization problem. It requires that the constraints of a given problem should be satisfied for all realizations of the uncertain parameters in a so-called uncertainty set. The robust version of a mathematical optimization problem is generally referred to as the *robust counterpart* (RC) problem. RO is popular because of the tractability of the RC for many classes of uncertainty sets, and its applicability in real life.

Adjustable robust optimization (ARO) distinguishes between two types of decision variables, namely, "wait and see" or "here and now". More precisely, decisions on "here and now" variables must be made before the uncertain parameters are realized like in classical RO; whereas decisions on "wait and see" variables are made after some or all of the uncertain parameters reveal themselves. ARO relates "wait and see" variables to uncertain parameters using decision rule functions; for additional details, we refer to (Ben-Tal et al., 2009, Chapter 14). In this chapter, we propose an ARO approach for a class of bilevel optimization problems. We assume that the first stage decisions are "here and now", and the later stage decisions are "wait and see", i.e., they are adjustable according to a realized portion of the uncertain data. It is important to point out that bilevel optimization problems that we shall consider in this chapter are in general computationally intractable, but by using modern RO techniques we provide practable approximations of the associated problems.

Bilevel optimization (BO) problems involve two nested traditional optimization problems with conflicting objectives. These two optimization problems are often referred to as the *upper-level* and the *lower-level* problems. The standard formulation is as follows

$$\min_{x\in\mathcal{X},\ y\in Y(x)} f(x,y) \quad \text{where} \quad Y(x) = \operatorname{argmin}_{z\in Z(x)}\ h(x,z), \tag{5.1}$$

where $x \in \mathcal{X} \subseteq \mathbb{R}^n$ is the upper-level decision, $y \in Y(x) \subseteq Z(x) \subseteq \mathbb{R}^m$ is the lower-level decision, $\mathcal{X}$ is the set of feasible solutions of the upper-level problem, $Z(x)$ is the set of feasible solutions of the lower-level problem that are dependent on the upper-level decision x, and $f, h : X \times Y \to \mathbb{R}$. Using the optimal value reformulation Dempe and Zemkoho (2013); Mitsos et al. (2008), (5.1) can be reformulated as the following single-level problem:

$$\min_{x\in\mathcal{X},\ y\in Z(x)} f(x,y) \ \text{ s.t. } \ h(x,y) \le h^*(x), \tag{5.2}$$

where $h^*(x) = \min\{h(x,y) \ : \ y \in Z(x)\}$. If the lower-level optimization problem is convex, then the most common approach is to replace this problem with the associated Karush-Kuhn-Tucker (KKT) conditions in order to obtain the explicit single-level reformulation. Nevertheless, the associated reformulation is often non-convex even for the linear case of the bilevel problem. Therefore, to solve these complex problems, global optimization algorithms are needed; for details of such algorithms, we refer to Faísca et al. (2007); Mitsos et al. (2008); Tsoukalas et al. (2009), and Tuy et al. (2007).

Even though bilevel optimization problems are computationally complex, they have many applications in real life. History of bilevel problems dates back to the 1930s, the seminal work of von Stackelberg proposes a bilevel problem to find the market equilibria in an economic model; see in Tirole and Fudenberg (1991). Forty years later, Bracken and McGill (1973) formalize bilevel problems as optimization problems. In the recent twenty years, such optimization problems have been applied to various fields, e.g., revenue management (Côté et al., 2003), supply chain management (Ryu et al., 2004), production planning (Iyer and Grossmann, 1998), transportation (Migdalas, 1995), security (Scaparra and Church, 2008), power markets (Zhang et al., 2011); and this list can be easily extended.

The *pessimistic* bilevel optimization problem that slightly deviates from the standard formulation (5.1) is defined as

$$\min_{x\in\mathcal{X}} \max_{y\in Y(x)} f(x,y) \quad \text{where} \quad Y(x) = \operatorname{argmin}_{z\in Z(x)} h(x,z), \tag{5.3}$$

where the problem is called *independent* when $Y(x) = Y(x')$ for all $x, x' \in \mathcal{X}$; and *dependent* when $Y(x) \neq Y(x')$ for some $x, x' \in \mathcal{X}$.

Note that the pessimistic problem (5.3) can be reformulated as

$$\begin{aligned} &\min_{x\in\mathcal{X},\tau} \tau \\ &\text{s.t. } \tau \geq f(x,y) \quad \forall y \in Y(x) = \operatorname{argmin}_{z\in Z(x)} h(x,z) \end{aligned} \tag{5.4}$$

using an epigraphic reformulation and "$\forall$" that replaces the inner maximization. Wiesemann et al. (2013) propose convergent approximation techniques to solve *independent* pessimistic optimization problems using *semi-infinite programming* under mild regularity assumptions.

It is important to point out that the pessimistic bilevel problem (5.3) is computationally more complex than the classical problem (5.1), and this is why most theoretical contributions to bilevel optimization are for the classical (*optimistic*) formulation that replaces "$\forall$" with "$\exists$" in (5.4). Both pessimistic and optimistic formulations have natural interpretations as non-cooperative games that are commonly referred to as the Stackelberg games between two players; for details again see Tirole and Fudenberg (1991). A Stackelberg game is a sequential game, where the first player (*the leader*) chooses his decision x, and then the second player (*the follower*) observes x and reacts with a decision y. In the pessimistic formulation, the leader has limited information about the optimal decision of the follower, and hence he wants to stay

feasible for any rational decision of the follower. On the other hand, in the optimistic formulation, the leader may bias the decisions of the follower to some extent and the follower altruistically favors the leader's objective with her decisions. The optimal solutions of the two formulations may significantly differ from each other.

In this chapter, we propose solution methods for a general pessimistic bilevel optimization problem with uncertainty. This problem is computationally intractable because the follower problem may be nonconvex, and the problem contains uncertain parameters as well as dependent second stage decisions. To solve the associated problem, we propose conservative approximation methods that are based on primal and dual linear decision rules and some other RO techniques that shall be discussed in later sections. The proposed technique simplifies the complex structure of the problem and provides a more tractable reformulation that gives a conservative solution. To quantify the optimality performance of this solution, we also give an optimistic reformulation that yields an efficient lower bound for the original problem. As well as being computationally less challenging, the associated reformulation method also has natural practical interpretations: first, the leader may have *incomplete information* about the follower decisions and this is why he may want to include some safety margin with a conservative formulation. Second, the follower may be constrained by *bounded rationality*, i.e., she may not solve the lower-level problem to optimality. In both cases the leader may want to hedge against the worst-case decisions the follower can make by solving a conservative reformulation, and our solution method also fits into this practical perspective.

The remainder of the chapter is organized as follows. §2 presents single-level reformulation methods for pessimistic bilevel optimization problems. §3 summarizes primal and dual linear decision rules used in RO. §4 proposes an RO methodology for a class of pessimistic bilevel optimization problems. §5 presents a numerical experiment. Finally, §6 summarizes our conclusions.

5.2 Single-Level Reformulation

In this section, we give a brief introduction on the reformulation techniques of pessimistic bilevel optimization problems with certain data. For the sake of exposition, we adopt an implicit notation in this section. Later in §§3 and 4, we shall explicitly dwell in details of these reformulation techniques.

To begin with, using the optimal value reformulation, the pessimistic formulation

(5.3) can be reformulated as

$$\begin{aligned} \min_{x\in\mathcal{X}} \ \max_{y\in Y(x)} \ & f(x,y) \\ \text{s.t.} \ & h(x,y) \le h^*(x), \end{aligned} \tag{5.5}$$

where $h^*(x) = \min\{h(x,y) \ : \ y \in Z(x)\}$ as in (5.2). Next, if the dual problem of the inner maximization problem in (5.5) satisfies the regularity conditions (in particular, it is a convex optimization problem), then it can be reformulated as

$$\max_{y\in Y(x)} \ f(x,y) \ \text{ s.t. } \ h(x,y) \le h^*(x) \quad = \quad \min_{\lambda} \ g(x,\lambda) \ \text{ s.t. } \ \lambda \in \Lambda(x), \tag{5.6}$$

where $g(x,\lambda)$ denotes the dual objective value function, $\lambda \in \Lambda(x)$ denotes the dual variables, and $\Lambda(x)$ is the set of dual feasible solutions for a given $x \in \mathcal{X}$. Eventually, the pessimistic bilevel problem (5.3) can be reformulated as the *single-level* problem:

$$\min_{\lambda,\ x\in\mathcal{X}} \ g(x,\lambda) \ \text{ s.t. } \ \lambda \in \Lambda(x). \tag{5.7}$$

To point out, going from (5.3) to (5.5) is justified because the regularity condition is assumed to be satisfied for the lower-level problem, i.e., we can use the dual problem to explicitly formulate $h^*(x)$. If the regularity condition is not met, then the lower-level problem can be replaced with KKT conditions but this yields a nonconvex and conservative reformulation of (5.3); see Mirrlees (1999). Similarly, going from (5.5) to (5.7) is justified only when the inner maximization problem (5.6) satisfies the constraint qualification, and it has a concave objective in y for any $x \in \mathcal{X}$ (e.g., when $f(x,y)$ is additively separable in x and y). Eventually, we obtain the single level reformulation in (5.7), but it is important to point out that the resulting reformulation is often a nonconvex optimization problem unless there exists a special convexification technique like in §4.

5.3 Primal and Dual Linear Decision Rules

In this section, we give a brief overview of the primal and dual linear decision rule methods used in stochastic and robust optimization. This chapter focuses on a specific class of the uncertain optimization problems where a decision maker first observes a portion of the uncertain parameter ξ and then selects a decision $y(\xi) \in \mathbb{R}^n$ that is subject to constraints $Ay(\xi) \le b(\xi)$ with a cost $c(\xi)^\top y(\xi)$. In SO, the goal is to find the optimal $y(\xi) \in \mathcal{L}^2_{k,n}$ that minimizes the expected cost, where $\mathcal{L}^2_{k,n}$ is the space of all Borel measurable, square-integrable functions from $\mathbb{R}^k$ to $\mathbb{R}^n$. Such

a decision $y(\xi)$ is often referred to as "wait and see"; and the decision problem that is mentioned above can be formulated as the following linear one-stage problem:

$$\begin{array}{ll} \min\limits_{y} & \mathbb{E}_{\xi}[c(\xi)^{\top} y(\xi)] \\ \text{s.t.} & Ay(\xi) \leq b(\xi) \quad \mathbb{P}_{\xi}\text{-a.s.}, \end{array} \tag{$\mathcal{SP}$}$$

where the coefficients matrix $A \in \mathbb{R}^{m \times n}$ is independent of ξ, i.e., the fixed recourse, and the constraints are satisfied almost surely with respect to $\mathbb{P}_{\xi}$, i.e., the underlying probability distribution of ξ. We assume that the objective function coefficients and the right-hand side are linear in ξ, more precisely, $c(\xi) = C\xi$ for some $C \in \mathbb{R}^{n \times k}$ and $b(\xi) = B\xi$ for some $B \in \mathbb{R}^{m \times k}$. In addition, we also assume that the support of the probability distribution $\mathbb{P}_{\xi}$ resides in a non-empty compact polyhedron:

$$\Xi = \left\{ \xi \in \mathbb{R}^k : W\xi \geq h \right\} \tag{5.8}$$

for some matrix $W \in \mathbb{R}^{l \times k}$ and a vector $h \in \mathbb{R}^l$. Without loss of generality, we assume that $\xi_1 = 1$ $(\forall \xi \in \Xi)$, i.e.,

$$W = (e_1, -e_1, \hat{W}^{\top})^{\top} \text{ and } h = (1, -1, 0, \ldots, 0)^{\top},$$

where e_1 denotes a basis vector in $\mathbb{R}^k$. We use $\xi_1 = 1$ to model the constant terms at affine functions.

It is known that optimizing $\mathcal{SP}$ is an NP-hard problem even when $\mathbb{P}_{\xi}$ is a uniform distribution on a unit cube; for details see Dyer and Stougie (2006, Theorem 3.2). Therefore, there is no polynomial time algorithm that solves this problem unless P = NP. Nevertheless Ben-Tal et al. (2004) propose a tractable approximation of $\mathcal{SP}$ by restricting the functional space of $y(\xi)$ to linear decision rules. It is important to point out that this complexity reduction method has been first proposed by Ben-Tal et al. in the realm of RO. Later Chen et al. (2008) and Shapiro and Nemirovski (2005) extended the ideas of Ben-Tal et al. to SO.

5.3.1 Primal Approximation

As it is mentioned above, a tractable approximation of $\mathcal{SP}$ is obtained by adopting linear decision rules, i.e., $y(\xi) = Y\xi$. The resulting problem contains a finite number of decision variables and infinitely many constraints. Therefore, it is a semi-infinite optimization problem given as follows

$$\begin{array}{ll} \min\limits_{Y \in \mathbb{R}^{n \times k}} & \operatorname{Tr}(MC^{\top}Y) \\ \text{s.t.} & AY\xi \leq B\xi \quad \mathbb{P}_{\xi}\text{-a.s.}, \end{array} \tag{5.9}$$

where $M := \mathbb{E}[\xi\xi^\top]$ is the second-order moment matrix. Notice that (5.9) is an upper bound for $\mathcal{SP}$ since it restricts the feasible functional space of the decision rules.

Tractable reformulation of (5.9). If we assume that the support of the probability distribution $\mathbb{P}_\xi$ resides in the compact polyhedron Ξ given in (5.8), then (5.9) can be equivalently reformulated as

$$\begin{array}{ll} \min\limits_{Y\in\mathbb{R}^{n\times k}} & \mathrm{Tr}(MC^\top Y) \\ \text{s.t.} & AY\xi \leq B\xi \quad \forall \xi \in \Xi \end{array} \tag{5.10}$$

Notice that (5.10) is also a semi-infinite optimization problem that seems to be intractable in its current form. However, it can be reformulated with a finite number of constraints by using the RO paradigm; for details on deriving a tractable robust reformulation, i.e., the *robust counterpart* (RC) problem, we refer to Ben-Tal and Nemirovski (1998), and Ben-Tal et al. (2009). Eventually (5.9) simplifies to the following linear optimization problem (LP):

$$\begin{array}{ll} \min & \mathrm{Tr}(MC^\top X) \\ \text{s.t.} & Y \in \mathbb{R}^{n\times k}, \Lambda \in \mathbb{R}^{m\times l}, \Lambda \geq 0 \\ & AY + \Lambda W = B \\ & \Lambda h \geq 0 \end{array} \tag{$\mathcal{SP}^u$}$$

where Λ is a matrix of auxiliary dual variables.

All in all, by adopting linear decision rules and using the RO paradigm, we obtain a tractable reformulation $\mathcal{SP}^u$ that constitutes an upper bound for $\mathcal{SP}$.

5.3.2 Dual Approximation

The dual problem of $\mathcal{SP}$, provided in Wright (1994), is given as follows

$$\inf_{y,s\geq 0} \sup_{\lambda} \mathbb{E}_\xi \left[c(\xi)^\top y(\xi) + \lambda(\xi)^\top [Ay(\xi) + s(\xi) - b(\xi)] \right] \tag{$\mathcal{SP}_d$}$$

where $s(\xi)$ denotes the decision rules for the slack variables, $\inf_{y,s\geq 0}$ denote the infimum over all $y \in \mathcal{L}^2_{k,n}$ and over all $s \in \mathcal{L}^2_{k,m}$ that are almost surely nonnegative. Similarly, $\sup_\lambda$ stands for the supremum over all $y \in \mathcal{L}^2_{k,m}$. Notice that the feasible functional spaces are kept general in $\mathcal{SP}_d$.

Similar to the complexity reduction techniques applied for the primal problem, Kuhn et al. (2011) propose adopting linear decision rules for the dual variables in $\mathcal{SP}_d$, more

precisely, $\lambda(\xi) = \Lambda\xi$, that is given as follows

$$\begin{aligned} &\inf_{y,s\geq 0} \sup_{\Lambda} \mathbb{E}_\xi \left[c(\xi)^\top y(\xi) + \xi^\top \Lambda^\top [Ay(\xi) + s(\xi) - b(\xi)]\right] \\ &= \inf_{y,s\geq 0} \sup_{\Lambda} \mathbb{E}_\xi \left[c(\xi)^\top y(\xi)\right] + \mathrm{Tr}\left(\Lambda^\top \mathbb{E}\left([Ay(\xi) + s(\xi) - b(\xi)]\xi^\top\right)\right). \end{aligned} \tag{5.11}$$

To point out, (5.11) constitutes a lower bound on the optimal objective value of $\mathcal{SP}$ since it restricts the feasible functional space of the dual variables to a linear decision rule in the dual problem. More precisely, the inner problem, i.e., the supremum, yields a lower optimal objective value (for a given y and s) compared to that in $\mathcal{SP}_d$ since the dual decision rules are restricted to linear ones. Notice that, in order (5.11) to be well defined, $\mathbb{E}\left([Ay(\xi) + s(\xi) - b(\xi)]\xi^\top\right) = 0$ must be satisfied, otherwise we can select Λ in such a way that the supremum goes to infinity and the problem becomes infeasible. Therefore, we can reformulate (5.11) as the following semi-infinite problem:

$$\begin{array}{ll} \min\limits_{y,s} & \mathbb{E}_\xi \left[c(\xi)^\top y(\xi)\right] \\ \text{s.t.} & \left.\begin{array}{l} \mathbb{E}_\xi \left[[Ay(\xi) + s(\xi) - b(\xi)]\xi^\top\right] = 0 \\ s(\xi) \geq 0 \end{array}\right\} \quad \mathbb{P}_\xi\text{-a.s. (or } \forall \xi \in \Xi). \end{array} \tag{5.12}$$

Tractable reformulation of (5.12). Similar to the primal approximation, using RO, Kuhn et al. (2011) propose an equivalent tractable reformulation of (5.12), when the support of $\mathbb{P}_\xi$ is in a compact polyhedron (5.8). The tractable RC reformulation is given as follows

$$\begin{array}{ll} \min & \mathrm{Tr}(MC^\top Y) \\ \text{s.t.} & Y \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{m\times k} \\ & AY + S = B \\ & (W - he_1^\top)MS^\top \geq 0, \end{array} \tag{$\mathcal{SP}^l$}$$

where M is the second-order moment matrix; for details on deriving $\mathcal{SP}^l$ we refer to Kuhn et al. (2011, §2.4).

Remark 12 *Notice that we can quantify the loss of optimality incurred by using linear decision rules in $\mathcal{SP}$ by adopting linear decision rules to the dual problem. In other words, $\mathcal{SP}^l$ constitutes a lower bound on $\mathcal{SP}^u$, and the gap between val($\mathcal{SP}^u$) and val($\mathcal{SP}^l$) estimates the approximation error of adopting linear decision rules.*

5.4 Robust Bilevel Optimization

In this section, we adopt primal and dual linear decision rules to approximate a class of bilevel optimization problems. The associated problem is a multi-stage uncertain optimization problem with two decision makers, namely, the *leader* and the *follower*. The decision making process is sequential, in the first stage, the leader makes binary $(0-1)$ decisions $x \in \mathcal{X} \subseteq \{0,1\}^q$ with a cost $q^\top x$ before the uncertain data $\xi \in \mathbb{R}^k$ reveals itself; at later stages, the follower observes an element ξ of the uncertain data and makes a decision $y'(\xi) \in \mathcal{L}^2_{k,n}$ with a cost $c(\xi)^\top y'(\xi)$ and subject to the constraints $Ay'(\xi) \leq b(x,\xi)$ that are dependent on the first stage decision x made by the leader. The leader knows the possible decisions $y(\xi)$ that the follower can select, and takes into account the worst-case cost $v(\xi)^\top y(\xi)$ that may be imposed on him by the follower. The associated decision problem can be formulated as the following bilevel optimization problem:

$$\begin{array}{lll} \min\limits_{x\in\mathcal{X}} & q^\top x + \max\limits_{y(.)} & \mathbb{E}_x i[v(\xi)^\top y(\xi)] \\ & \text{s.t.} & y(\cdot) \in \operatorname{argmin}_{y'(.)} \left\{ \mathbb{E}_\xi[c(\xi)^\top y'(\xi)] : Ay'(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi \right\}, \end{array} \tag{$\mathcal{BP}$}$$

where Ξ denotes the non-empty compact polyhedron (5.8) that contains the support of the probability distribution $\mathbb{P}_\xi$ of ξ.

Remark 13 *If $y(\cdot)$ is a continuous function and the support of $\mathbb{P}_\xi$ is contained in the compact set Ξ, then $[Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi \equiv Ay(\xi) \leq b(x,\xi) \quad \mathbb{P}_\xi$-a.s.] is satisfied.*

Some important properties of $\mathcal{BP}$ are: 1) the inner maximization problem of the upper-level optimization problem denotes the worst-case expected cost that shall be imposed on the leader by the follower's future decisions, i.e., we see an optimistic follower but a pessimistic leader due to different levels of information. 2) The leader and the follower have different cost functions, i.e., $v(\xi) \neq c(\xi)$; the special case $v(\xi) = c(\xi)$ will be treated later in Theorem 6. 3) The constraints in the lower-level optimization problem ensures that the follower's decisions $y'(\xi)$ are robust against all possible realizations of the uncertain problem data, more precisely, $Ay'(\xi) \leq b(x,\xi)$ for all $\xi \in \Xi$. 4) The leader decides on x before the uncertainty is realized, i.e., x is "here and now"; the follower decides on $y(\xi)$ after the uncertain data ξ reveals itself, i.e., $y(\xi)$ is "wait and see". 5) We assume $c(\xi)$ and $v(\xi)$ are linear in ξ, i.e., $c(\xi) = C\xi$ for some $C \in \mathbb{R}^{n\times k}$ and $v(\xi) = V\xi$ for some $V \in \mathbb{R}^{n\times k}$; and $b(x,\xi)$ is linear in x and ξ independently (so interactions are allowed), i.e., $b(x,\xi) = B_x\xi$ where $B_x \in \mathbb{R}^{m\times k}$ denotes the variable matrix such that each component $B_{ij}(x)$ is a linear function of

x, which is a vector of $0-1$ variables.

Note that $\mathcal{BP}$ is a complex semi-infinite optimization problem that is intractable, nevertheless, using RO techniques and primal and dual linear decision rules we can derive a practable approximation of $\mathcal{BP}$. Theorem 6 proposes such an approximation of $\mathcal{BP}$.

Theorem 5 *The following formulation gives a conservative approximation of $\mathcal{BP}$ by adopting primal and dual linear decision rules.*

$$\begin{array}{lll} \min\limits_{x\in\mathcal{X},\,\Lambda_{1,\ldots,4},\,\lambda} & q^\top x + \mathrm{Tr}\left(\Lambda_1 B_x - \Lambda_3 M C^\top\right) & \\ s.t. & -MV^\top + \lambda M C^\top + \Lambda_1 A & = 0 \\ & -\lambda B_x - A\Lambda_3 + \Lambda_4 W & = 0, \\ & \Lambda_1^\top + \Lambda_2(he^\top - W)M & = 0 \\ & \Lambda_4 h, \Lambda_2, \Lambda_4, \lambda & \geq 0 \end{array} \qquad (\mathcal{BP}^u)$$

where $x \in \mathcal{X} \subseteq \{0,1\}^g$; $\Lambda_1 \in \mathbb{R}^{k\times k}, \Lambda_2 \in \mathbb{R}^{k\times l}, \Lambda_3 \in \mathbb{R}^{n\times k}, \Lambda_4 \in \mathbb{R}^{m\times l}$, and $\lambda \in \mathbb{R}$ are the auxiliary dual variables.

Proof. In the remainder, we focus on the three steps to derive the associated conservative RC approximation $\mathcal{BP}^u$.

Step 1 (*Adopting primal linear decision rules*). The bilevel problem $\mathcal{BP}$ can be transformed into an equivalent single-level problem as follows

$$\begin{array}{lll} \min\limits_{x\in\mathcal{X}} \quad q^\top x + \max\limits_{y(\cdot)} & \mathbb{E}_\xi[v(\xi)^\top y(\xi)] & \\ \text{s.t.} & \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \leq z^*(x) & \\ & Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi, & \end{array} \qquad (5.13)$$

where $z^*(x)$ denotes the optimal objective value of the lower-level (or the follower) problem for a given x:

$$\begin{array}{lll} z^*(x) = \min\limits_{y(\cdot)} & \mathbb{E}_\xi[c(\xi)^\top y(\xi)] & \\ \text{s.t.} & Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi. & \end{array} \qquad (5.14)$$

As we have explained earlier in §5.3, problem (5.14) is intractable, however, we can derive a more tractable approximation ($\mathcal{SP}^u$) by adopting linear decision rules to

$y(\xi)$. Notice that the optimal objective value of $\mathcal{SP}^u$ yields an upper bound to that of (5.14), and so does the optimal objective value of the dual problem of $\mathcal{SP}^u$:

$$\begin{aligned} \min_{\Gamma,\theta\geq 0} \quad & \mathrm{Tr}(\Gamma B_x) && (d\mathcal{SP}^u)\\ \text{s.t.} \quad & MC^\top - \Gamma A = 0 && (5.15)\\ & W\Gamma + h\theta^\top \geq 0, && (5.16) \end{aligned}$$

where $\Gamma \in \mathbb{R}^{l\times m}$ and $\theta \in \mathbb{R}^l$ denote the auxiliary dual variables, and $B_x \in \mathbb{R}^{m\times k}$ comes from $b(x,\xi) = B_x\xi$.

To point out, there is no optimality gap between val($\mathcal{SP}^u$) and its dual val($d\mathcal{SP}^u$) since $\mathcal{SP}^u$ is an LP when x is fixed, i.e., val($\mathcal{SP}^u$)=val($d\mathcal{SP}^u$) at optimality. Therefore, we can replace $z^*(x)$ in (5.13) with the objective value of $d\mathcal{SP}^u$, moreover, we can delete the minimization term in the dual objective as its often done in RO, since the dual problem yields an upper bound for the original problem. Then, the problem becomes

$$\begin{aligned} (\leq)\ \min_{x\in\mathcal{X}} \quad & q^\top x + \max_{y(\cdot),\Gamma,\theta\geq 0} && \mathbb{E}_\xi[v(\xi)^\top y(\xi)] && (5.17)\\ & \text{s.t.} && \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \leq \mathrm{Tr}(\Gamma B_x) && (5.18)\\ & && Ay(\xi) - b(x,\xi) \leq 0 \quad \forall \xi \in \Xi && (5.19)\\ & && MC^\top - \Gamma A = 0 && (5.20)\\ & && W\Gamma + h\theta^\top \geq 0, && (5.21) \end{aligned}$$

where (5.20) and (5.21) represent the additional constraints that come from the dual problem $d\mathcal{SP}^u$. To sum up, we use the primal approximation of (5.14) to relax the right-hand side of the first constraint in (5.13), i.e., $\mathbb{E}_\xi[c(\xi)^\top y(\xi)] \leq z^*(x)$. The optimal objective value of the new reformulation yields an upper bound for $\mathcal{BP}$, and the associated relation is presented by "$(\leq)$" before the formulation in (5.17). Notice that we do not impose any restrictive assumption on the feasible functional space of $y(\xi)$ in (5.17)$-$(5.19).

Step 2 (*Adopting dual linear decision rules*). Now using the dual approximation technique that is described in §5.3.2, we will relax the intractable left-hand side of (5.19). First, we relax the semi-infinite constraint (5.19) as $\mathbb{E}_\xi\left[[Ay(\xi) + s(\xi) - b(\xi)]\xi^\top\right] = 0$;

see (5.12) in §5.3.2, and the new formulation becomes

$$
\begin{array}{lll}
(\leq\leq)\ \min\limits_{x\in\mathcal{X}} & q^\top x + \max\limits_{y(\cdot),\Gamma,\theta\geq 0} & \mathbb{E}_\xi[\xi^\top V^\top y(\xi)] \\
& \text{s.t.} & \mathbb{E}_\xi[\xi^\top C^\top y(\xi)] \leq \mathrm{Tr}(\Gamma B_x) \\
& & \mathbb{E}_\xi\left[[Ay(\xi) + s(\xi) - B_x\xi]\xi^\top\right] = 0 \\
& & s(\xi) \geq 0 \quad \forall \xi \in \Xi \\
& & MC^\top - \Gamma A = 0 \\
& & W\Gamma + h\theta^\top \geq 0.
\end{array}
\tag{5.22}
$$

It can be shown that the original decisions $y(\xi) \in \mathcal{L}^2_{k,n}$ and $s(\xi) \in \mathcal{L}^2_{k,m}$ can be determined as $YM = \mathbb{E}_\xi[y(\xi)\xi^\top]$ and $SM = \mathbb{E}_\xi[s(\xi)\xi^\top]$, $s(\xi) \geq 0\ \forall \xi \in \Xi$, where $Y \in \mathbb{R}^{n\times k}$, $S \in \mathbb{R}^{m\times k}$, and $M = \mathbb{E}_\xi[\xi\xi^\top]$; we refer to Kuhn et al. (2011, Proposition 2) for the formal proof of $[SM = \mathbb{E}_\xi[s(\xi)\xi^\top]]$. Then, we derive the RC of the second and the third constraints in (5.22) as in §5.3.2. Consequently, the approximation reduces to

$$
\begin{array}{lll}
(\leq\leq)\ \min\limits_{x\in\mathcal{X}} & q^\top x + \max\limits_{Y,S,\Gamma,\theta\geq 0} & \mathrm{Tr}(V^\top YM) \\
& \text{s.t.} & \mathrm{Tr}(C^\top YM) - \mathrm{Tr}(\Gamma B_x) \leq 0 \\
& & AYM + SM = B_xM \\
& & (W - he_1^\top)MS^\top \geq 0 \\
& & MC^\top - \Gamma A = 0 \\
& & W\Gamma + h\theta^\top \geq 0,
\end{array}
\tag{5.23}
$$

where $\theta \in \mathbb{R}^l$, and $\Gamma \in \mathbb{R}^{m\times k}$.

As it is explained above, we relax the semi-infinite constraints in (5.19) by adopting the dual approximation techniques, therefore, the optimal objective value of (5.23) yields an upper bound for (5.17). We present the associated relation by "$(\leq\leq)$".

Remark 14 *Notice that M is an invertible matrix. Therefore, we can take the product of both sides of the equality constraint $[AYM + SM = B_xM]$ in* (5.23) *by M^{-1}, and it reduces to $AY + S = B_x$.*

Step 3 (*Duality*). To simplify the minimax formulation, we take the dual of the inner maximization problem in (5.23). Eventually, it gives the "final" approximation $\mathcal{BP}^u$. To point out, $\mathcal{BP}^u$ contains bilinear terms in the objective function and in the second constraint (i.e., $\Lambda_1 B_x$ and $\lambda B_x M$); however, such expressions can be linearized, and the exact linear reformulations of such bilinear expressions are treated below in Proposition 1 and 2. ■

Linearization. Notice that the components of the matrix $B_x \in \mathbb{R}^{m\times k}$ are linear functions of binary variables $x \in \mathcal{X} \subseteq \{0,1\}^q$, therefore, the bilinear expressions in $\mathcal{BP}^u$ contain first order interactions between the binary variable x and the continuous variables λ, and Λ_1. For the sake of exposition, we focus on (say) a binary $(0-1)$ variable x and a continuous variable λ, and (say) the bilinear term is $\alpha\lambda x$ where α denotes a constant. If we define a new continuous variable w such that $w := \alpha\lambda x$, then we can linearize the bilinear expression by using a group of additional constraints and the variable w, and the associated reformulations are exact. Proposition 1 presents a method to linearize such bilinear terms in the constraints; and Proposition 2 shows the same for the objective function.

Proposition 1 *A bilinear expression $\alpha\lambda x$ of variables $\lambda \in \mathbb{R}^+$ and $x \in \{0,1\}$ (with a constant $\alpha \in \mathbb{R}$) can be linearly expressed by the following group of constraints:*

$$w \leq \alpha\lambda, \quad w \leq \alpha\lambda^u x, \quad w \geq \alpha(\lambda - \lambda^u(1-x)),$$

where $w = \alpha\lambda x$, and λ^u denotes an upper bound for λ (i.e., $0 \leq \lambda \leq \lambda^u$).

Remark 15 *Notice that the continuous variable λ in $\mathcal{BP}^u$ is unrestricted. However, we can find an efficient upper bound λ^u by setting an initial upper bound, and then iteratively re-optimizing the problem at hand for increasing value of λ^u as long as the optimal solution does not change.*

Proposition 2 *A bilinear expression in an objective function such as*

$$\min_{x\in\{0,1\};\ \gamma\in\mathbb{R}} \alpha\gamma x$$

can be linearized as

$$\min_{x\in\{0,1\};\ w,\gamma\in\mathbb{R}} w \quad s.t. \quad \alpha\gamma \leq w + Q(1-x), \quad 0 \leq w + Qx$$

where $w = \alpha\lambda x$, and Q denotes a large constant.

Therefore, we can conclude that using Proposition 1 and 2, $\mathcal{BP}^u$ can be reformulated as a mixed integer LP (MILP).

Theorem 6 *If $c(\xi) = v(\xi)$, then the bilevel optimization problem $\mathcal{BP}$ yields the same optimal objective value as the following single-level problem*

$$\begin{aligned} \min_{x\in\mathcal{X},y} \quad & q^\top x + \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \\ s.t. \quad & Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi. \end{aligned} \tag{1$\mathcal{P}$}$$

Proof. Let $Y(x)$ denote the optimal functional space of the lower-level (follower) problem in $\mathcal{BP}$ for a given first stage decision $x \in \mathcal{X}$, and suppose $c(\xi) = v(\xi)$, i.e,

$$\mathcal{BP} := \min_{x \in \mathcal{X}} \; q^\top x + \max_{y(\cdot) \in Y(x)} \mathbb{E}_\xi[c(\xi)^\top y(\xi)],$$

where $Y(x) = \text{argmin}_{y'(.)} \left\{ \mathbb{E}_\xi[c(\xi)^T y'(\xi)] : Ay'(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi \right\}$. Now, we shall consider the two possible cases of $Y(x)$, more precisely, when it has a unique solution, and when it has alternative optimal solutions.

• If $Y(x)$ is unique for any given $x \in \mathcal{X}$, then we can replace the inner maximization problem in $\mathcal{BP}$ with the minimization problem, more precisely,

$$\max_{y \in Y(x)} \; \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \;\equiv\; \min_{y(\cdot)} \; \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \;\text{ s.t. }\; Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi. \tag{5.24}$$

Therefore, $\mathcal{BP}$ becomes

$$\min_{x \in \mathcal{X}, y(\cdot)} \; q^\top x + \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \;\text{ s.t. }\; Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi \quad (\equiv 1\mathcal{P}).$$

• If $Y(x)$ is not unique for some $x \in \mathcal{X}$, then (5.24) still holds since the alternative optimal solutions in $Y(x)$ yield the same objective value by definition. ■

Corollary 2 *The conservative approximation of* $1\mathcal{P}$*:*

$$\min_{x \in \mathcal{X}, y(\cdot)} \; q^\top x + \text{Tr}(C^\top Y M) \;\; \textit{s.t.} \;\; AY\xi \leq B_x \xi \quad \forall \xi \in \Xi, \tag{$1\mathcal{P}^u$}$$

i.e., obtained by adopting linear decision rules to $y(\xi)$ *in* $1\mathcal{P}$*, yields the same optimal objective value with the conservative approximation* $\mathcal{BP}^u$ *when* $c(\xi) = v(\xi)$*.*

Proof. The inner maximization problem $[\max_{y(.)} \; \mathbb{E}_\xi[v(\xi)^\top y(\xi)]]$ in $\mathcal{BP}$ is bounded above by the dual objective value in the original approximation when $c(\xi) = v(\xi)$; see (5.18). This dual objective value constitutes a lower bound for the follower problem, i.e., $[\min_{y(.)} \; \mathbb{E}_\xi[c(\xi)^\top y(\xi)] \;:\; Ay(\xi) \leq b(x,\xi) \quad \forall \xi \in \Xi]$). Consequently, $\text{val}(\mathcal{BP}^u) = \text{val}(1\mathcal{P}^u)$. ■

Estimating the optimality performance of $\mathcal{BP}^u$. So far we have shown the derivation of the conservative approximation of $\mathcal{BP}$ by adopting primal and linear decision rules. Now we shall propose a progressive approximation of an optimistic reformulation problem. Consequently, this progressive approximation may be used

to quantify the optimality performance of the conservative approximation. The optimistic reformulation of the bilevel optimization problem is given as follows

$$\begin{aligned} \min_{x\in\mathcal{X}} \quad & q^\top x + \min_{y(\cdot)} \quad \mathbb{E}_\xi[v(\xi)^\top y(\xi)] \\ & \text{s.t.} \quad y(\cdot) \in \operatorname{argmin}_{y'(\cdot)} \left\{ \mathbb{E}_\xi[c(\xi)^T y'(\xi)] : Ay'(\xi) \le b(x,\xi) \quad \forall \xi \in \Xi \right\}. \end{aligned} \tag{$\mathcal{BP}_o$}$$

Different than $\mathcal{BP}$, $\mathcal{BP}_o$ models an optimistic leader objective, more precisely, the leader optimizes with respect to the best follower decision $y(\xi)$ that minimize his cost function $\mathbb{E}_\xi[v(\xi)^\top y(\xi)]$.

Even though $\mathcal{BP}_o$ is computationally less challenging than the original problem $\mathcal{BP}$, it is still an NP-Hard problem. This is why in Corollary 3, we provide a progressive approximation of this problem using primal and dual decision rules.

Corollary 3 *The following formulation gives a progressive approximation of the optimistic reformulation $\mathcal{BP}_o$ by adopting primal and dual linear decision rules.*

$$\begin{aligned} \min_{x\in\mathcal{X},Y,S,\Gamma,\theta\ge 0} \quad & q^\top x + \operatorname{Tr}(V^\top Y M) \\ s.t. \quad & \operatorname{Tr}(C^\top Y M) - \operatorname{Tr}(B_x^\top \Gamma) \le 0 \\ & AY + S = B_x \\ & (W - he_1^\top) M S^\top \ge 0 \\ & MC^\top - \Gamma A = 0 \\ & W\Gamma + h\theta^\top \ge 0, \end{aligned} \tag{$\mathcal{BP}_o^l$}$$

where $x \in \mathcal{X} \subseteq [0,1]^g, Y \in \mathbb{R}^{n\times k}, S \in \mathbb{R}^{m\times k}, \theta \in \mathbb{R}^l$, and $\Gamma \in \mathbb{R}^{m\times k}$.

Proof. Follows similar to Step 1 & 2 of the proof of Theorem 5. ■

Different than Theorem 5, in Corollary 3 we found a progressive bound, i.e., a lower bound, for the leader's objective at $\mathcal{BP}_o$. This is because we relax the problem two times by adopting primal and dual linear decision rules at Step 1 and Step 2; respectively, and the leader's objective is a minimization. Notice that the optimal objective value of the leader in the optimistic reformulation $\mathcal{BP}_o$ yields a lower bound for that in the conservative formulation $\mathcal{BP}$. Therefore, the optimal objective value of the progressive approximation $\mathcal{BP}_o^l$ also yields a lower bound for that in $\mathcal{BP}$, i.e., $\text{val}(\mathcal{BP}_o^l) \le \text{val}(\mathcal{BP}) \le \text{val}(\mathcal{BP}^u)$. Eventually, we may use the progressive approximation $\mathcal{BP}_o^l$ to "safely" estimate the optimality performance of the conservative approximation $\mathcal{BP}^u$.

5.5 Conclusion

In this chapter, we have proposed conservative and progressive approximation methods for a class of pessimistic bilevel optimization problems with uncertain data. We develop an RO approach using primal and dual decision rule reformulations. Even though our method uses linear decision rules to reduce the complexity of the original formulation, it does not restict the feasible function space of the adjustable decision variables. Moreover, we can estimate the approximation error of the proposed reformulation. To the best of our knowledge, this is the first publication that offers robust reformulations of pessimistic bilevel optimization problems. It is important to point out that using our RO method, the original formulation that is strictly NP-hard is reduced into a substantially easier MILP class. The associated approximation can be solved for moderate sized instances using commercial solvers, nevertheless, its performance is still highly dependent on the number of binary variables in the original problem. In future research, we shall investigate managerial implications of our method in real life problems.

Bibliography

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805.

Bracken, J. and McGill, J. T. (1973). Mathematical programs with optimization problems in the constraints. *Operations Research*, 21(1):37–44.

Chen, X., Sim, M., Sun, P., and Zhang, J. (2008). A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357.

Côté, J.-P., Marcotte, P., and Savard, G. (2003). A bilevel modelling approach to pricing and fare optimisation in the airline industry. *Journal of Revenue and Pricing Management*, 2(1):23–36.

Dempe, S. and Zemkoho, A. (2013). The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. *Mathematical Programming*, 138(1-2):447–473.

Dyer, M. and Stougie, L. (2006). Computational complexity of stochastic programming problems. *Mathematical Programming*, 106(3):423–432.

Faísca, N. P., Dua, V., Rustem, B., Saraiva, P. M., and Pistikopoulos, E. N. (2007). Parametric global optimisation for bilevel programming. *Journal of Global Optimization*, 38(4):609–623.

Iyer, R. R. and Grossmann, I. E. (1998). A bilevel decomposition algorithm for long-range planning of process networks. *Industrial & Engineering Chemistry Research*, 37(2):474–481.

Kuhn, D., Wiesemann, W., and Georghiou, A. (2011). Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209.

Migdalas, A. (1995). Bilevel programming in traffic planning: models, methods and challenge. *Journal of Global Optimization*, 7(4):381–405.

Mirrlees, J. A. (1999). The theory of moral hazard and unobservable behaviour: Part i. *The Review of Economic Studies*, 66(1):3–21.

Mitsos, A., Lemonidis, P., and Barton, P. (2008). Global solution of bilevel programs with a nonconvex inner program. *Journal of Global Optimization*, 42(4):475–513.

Ryu, J.-H., Dua, V., and Pistikopoulos, E. N. (2004). A bilevel programming framework for enterprise-wide process networks under uncertainty. *Computers & Chemical Engineering*, 28(6):1121–1129.

Sashihara, S. (2011). *The Optimization Edge: Reinventing Decision Making to Maximize All Your Company's Assets*. McGraw-Hill.

Scaparra, M. P. and Church, R. L. (2008). A bilevel mixed-integer program for critical infrastructure protection planning. *Computers & Operations Research*, 35(6):1905–1923.

Shapiro, A. and Nemirovski, A. (2005). On complexity of stochastic programming problems. In *Continuous optimization*, pages 111–146. Springer.

Tirole, J. and Fudenberg, D. (1991). *Game Theory*. MIT Press, Cambridge, MA.

Tsoukalas, A., Rustem, B., and Pistikopoulos, E. N. (2009). A global optimization algorithm for generalized semi-infinite, continuous minimax with coupled constraints and bi-level problems. *Journal of Global Optimization*, 44(2):235–250.

Tuy, H., Migdalas, A., and Hoai-Phuong, N. (2007). A novel approach to bilevel nonlinear programming. *Journal of Global Optimization*, 38(4):527–554.

Wiesemann, W., Tsoukalas, A., Kleniati, P., and Rustem, B. (2013). Pessimistic bilevel optimization. *SIAM Journal on Optimization*, 23(1):353–380.

Wright, S. (1994). Primal-dual aggregation and disaggregation for stochastic linear programs. *Mathematics of Operations Research*, 19(4):893–908.

Zhang, G., Zhang, G., Gao, Y., and Lu, J. (2011). Competitive strategic bidding optimization in electricity markets using bilevel programming and swarm technique. *Industrial Electronics, IEEE Transactions on*, 58(6):2138–2146.