Distributionally and integer adjustable robust optimization

Postek, Krzysztof

Publication date:
2017

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

Distributionally and integer adjustable robust optimization

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op

woensdag 15 februari 2017 om 16.00 uur
door

Krzysztof Stanisław Postek

geboren op 5 maart 1989 te Bytom, Polen

Promotiecommissie:

Promotores:	prof.dr.ir. Dick den Hertog
	prof.dr. Bertrand Melenberg
	prof.dr. Aharon Ben-Tal
Overige leden:	prof.dr. Etienne de Klerk
	prof.dr. Daniel Kuhn
	dr. Marjan van den Akker
	dr. Ruud Brekelmans

Distributionally and integer adjustable robust optimization

Acknowledgements

I would like to thank my supervisors Dick den Hertog and Bertrand Melenberg with whom my cooperation started five years ago, when I was still a Master student in Quantitative Finance and Actuarial Science. They have been extremely important people to my — I allow myself to use this word — growth throughout this period, both in terms of research and as a person.

Special thanks go to Aharon Ben-Tal for giving me the opportunity to visit the Technion for three months. This stay has shown me a different and challenging way of working on Mathematics, beginning a cooperation that continues and has resulted in my come-back to Haifa as a postdoctoral fellow.

Also, I would like to thank Maarten van der Vlerk and Ward Romeijnders for hosting me for a talk at the Workshop Stochastic Optimization in Groningen in September 2015 that resulted in combining our research in a fruitful and promising project. It is with a great shock that I have received the news about Maarten's death in October 2016 and a great pity that I cannot invite him to my public defense.

I would also like to thank two other people without whom my adventure with OR would not have started: Juan Vera, whose course on 'Nonlinear and Robust Optimization' taught me about this field, and Etienne de Klerk, who encouraged me to enter the Research Master in OR.

Special thanks go all the RM/PhD fellows from the K building. In particular, to those with whom I share the fate since 2011/2012: Marleen Balvert, Maria Lavrutich, Marieke Musegaas, and Trevor Zhen. Going outside the K building, I would like to thank three other people whose presence made this stay great and inspiring: Yana Yarovikova, Branislav Hock and Sr. Miriam Venner. Also, life in Tilburg during the last four out of my five years there would have been completely different without the Student Parish 'Maranatha' under the guidance of Fr. Michiel Peeters.

In the end, I am thankful to my family without whose home education and support none of these things would have ever been possible.

Krzysztof Postek
Tilburg, December 2016

Contents

CHAPTER 1

Introduction

1.1 Robust Optimization

A classical mathematical optimization problem is formulated as:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & f_0(\boldsymbol{x}, \boldsymbol{z}) \\ \text{s.t.} & f_i(\boldsymbol{x}, \boldsymbol{z}) \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{1.1}$$

where $\boldsymbol{x} \in \mathbb{R}^{n_1}$ is the decision vector, $\boldsymbol{z}$ is a problem parameter, and $f_i(\cdot, \cdot)$, $i = 0, \ldots, m$ are real-valued functions. In such a problem, one minimizes the value of the objective function subject to a number of constraints, parameterized by $\boldsymbol{z}$, which may involve parameters of the problem corresponding to a variety of possibilities: precision of devices (for example, the antenna experiment of Chapter 2), customer demand forecasts (for example, the lot sizing experiment of Chapter 3), lengths of procedures (for example, the operating room experiment of Chapter 5), climate changes (for example, the flood risk management experiment of Chapter 6), vehicle speeds, concentration of substances, to name only a few. All such values are subject to uncertainty resulting from

- measurement errors: imprecision of the physical measurement devices such as thermometers or missing inventory data;
- forecast errors: lack of precise knowledge about future product demand, discount rates, or resource prices;
- implementation errors: imprecision of implementation of devices such as electrical power or metal component lengths;
- reliability of past data (known also as veracity in the Big Data context): wrongly entered data, missing observations, and other data unreliabilities.

Even very small deviations from assumed parameter values can lead to huge infeasibilities of the problem constraints or to large losses in the objective function value,

see Ben-Tal et al. (2009). This fact calls for optimization methods able to include and accommodate for uncertainty. The growing popularity of modelling with explicit inclusion of uncertainty is illustrated by the number of search hits for 'robust optimization' in the Google Scholar[1] database from the subsequent time periods:

Time period	**Number of publications**
1990-1995	314
1995-2000	1110
2000-2005	2810
2005-2010	8100
2010-2015	15600

Robust Optimization (RO) is a modeling paradigm where the way of handling uncertainty is to construct an uncertainty set $\mathcal{Z}$ of values of $\boldsymbol{z}$ and to require the problem's constraints to hold for all $\boldsymbol{z} \in \mathcal{Z}$, and to assume the worst-possible outcome also for the objective function. The optimization problem is then formulated as:

$$\begin{array}{lll} \min\limits_{\boldsymbol{x},t} & t & \\ \text{s.t.} & f_0(\boldsymbol{x},\boldsymbol{z}) \leq t, & \forall \boldsymbol{z} \in \mathcal{Z} \\ & f_i(\boldsymbol{x},\boldsymbol{z}) \leq 0, & \forall \boldsymbol{z} \in \mathcal{Z} \quad i = 1,\ldots,m. \end{array} \tag{1.2}$$

Research in RO is primarily concerned with identifying the properties of

- functions $f(\cdot,\cdot)$ (convex, concave, piecewise linear,...),
- types of parameters $\boldsymbol{z}$ (scalars, vectors, probability distributions),
- types of uncertainty sets $\mathcal{Z}$ (box, ellipsoidal, semidefinite-representable...),

such that (1.2) can be reformulated to an equivalent (or approximate) *computationally tractable* form without the 'for all' condition, that is, one that can be solved in an *efficient* way. The word 'efficient' stands most often for the solution time being polynomial in the size of the problem input for continuous problems or a 'reasonable' solution time for problems with integer variables or nonconvex problems. In this work, RO is understood as reformulating a constraint

$$f_i(\boldsymbol{x},\boldsymbol{z}) \leq 0, \quad \forall \boldsymbol{z} \in \mathcal{Z} \tag{1.3}$$

by means of *duality* to an equivalent system of deterministic constraints (without the '$\forall$' symbol):

$$\begin{cases} G_i(\boldsymbol{x},\boldsymbol{\theta}_i) \leq 0 \\ \boldsymbol{\theta}_i \in \Theta_i(\boldsymbol{x}), \end{cases} \tag{1.4}$$

[1] As for April 21, 2016.

where $\boldsymbol{\theta}_i$ is a potential extra variable (where it can be the case that no extra variables are needed), $G_i(\cdot,\cdot)$ is a vector-valued function, and $\Theta_i(\boldsymbol{x})$ is an $\boldsymbol{x}$-dependent set. The following example illustrates this notion.

Example 1.1 *Suppose the constraint is:*

$$(\boldsymbol{a}+\boldsymbol{P}\boldsymbol{z})^T\boldsymbol{x}\leq 0,\quad \forall \boldsymbol{z}:\|\boldsymbol{z}\|_\infty\leq 1, \tag{1.5}$$

where $\boldsymbol{a},\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{z}\in\mathbb{R}^m$, and $\boldsymbol{P}\in\mathbb{R}^{n\times m}$. Then, since the dual norm of the ∞-norm is the 1-norm, we have that

$$\sup_{\boldsymbol{z}:\|\boldsymbol{z}\|_\infty\leq 1}\boldsymbol{z}^T\boldsymbol{P}^T\boldsymbol{x}=\|\boldsymbol{P}^T\boldsymbol{x}\|_1,$$

which implies that (1.5) is equivalent to the following deterministic constraint:

$$\boldsymbol{a}^T\boldsymbol{x}+\|\boldsymbol{P}^T\boldsymbol{x}\|_1\leq 0,$$

which can be reformulated as a system of linear constraints whose number grows linearly with m and n.

It should be also noted that another approach is possible, where one begins to solve a problem with only one scenario $\boldsymbol{z}^1\in\mathcal{Z}$. Then, the optimization problem is solved so that the solution is robust to it. After the optimal solution has been found, another scenario $\boldsymbol{z}^2\in\mathcal{Z}$ is determined that violates one or more constraints. In the next round, a solution is found that is robust to $\{\boldsymbol{z}^1,\boldsymbol{z}^2\}$ and the procedure is repeated, until the solution is robust to all possible scenarios. Such an iterative procedure is known as the *adversarial approach* (Bienstock 2007), and even though in certain applications it might be even more efficient than the reformulation approach, it is not considered in this thesis.

Having its roots in the paper by Soyster (1973), RO has experienced a dynamic growth since the late 1990s, after release of the works showing the impact of data uncertainty and providing tractable reformulations of RO problems: Ben-Tal and Nemirovski (1998, 2000, 2002) and El Ghaoui and Lebret (1997), El Ghaoui et al. (1998). The notable contributions of the 'early period' of RO (till around 2010) are Goldfarb and Iyengar (2003), Bertsimas and Sim (2004), Bertsimas et al. (2004), Calafiore and Campi (2005), Chen et al. (2007). In these papers, robust reformulations are constructed for a number of practically relevant settings.

Since around 2010 till now, next to problem-specific contributions to the field, authors started to provide unifying treatments of the results obtained up to now. The first and, up to now, only book in the field of RO - the monograph of Ben-Tal et al. (2009) - gives a unifying treatment of RO problems in terms of conic duality. Ben-Tal et al. (2015) provide another way of reformulating robust constraints based on Fenchel duality. Thanks to these contributions we know that problem (1.2) is tractable, for example, when:

- the constraints are linear and the uncertainty sets is an intersection of the standard, Lorentz, or the positive semidefinite cones (Ben-Tal and Nemirovski 1998),

- the constraints are second-order conic and the uncertainty set is an ellipsoid (El Ghaoui and Lebret 1997),

- the functions $f(\boldsymbol{x}, \cdot)$ are concave, and the uncertainty set is described by a finite set of convex constraints (Ben-Tal et al. 2015).

Next to these developments, there are developments in the field of polynomial optimization (Lasserre 2009), where duality for moment problems is used to construct either direct reformulations or converging hierarchies of approximations to the case where $f(\boldsymbol{x}, \boldsymbol{z})$ is a polynomial and the $\mathcal{Z}$ is semi-algebraic, see, for example, the paper Laraki and Lasserre (2008) or the recent book Lasserre (2015).

A question that still remains to be answered is the following: For what types of constraints on risk measures can we use the existing derivation techniques to construct a generic framework to obtain their tractable counterparts under distributional ambiguity?

In Chapter 2 we show how Fenchel duality techniques can be utilized in derivations of robust counterparts for constraints on risk measures of decision-dependent random variables, which is relevant to fields such as economics, finance, machine learning, and engineering. We also show in Chapter 2 how computationally tractable robust counterparts can be derived for nonlinear risk measures (e.g., variance) of decision-dependent random variables by exploiting their suitable formulations.

Since its founding, RO has been successfully applied to fields such as: inventory management (Ben-Tal et al. 2004), facility location (Ordonez and Zhao 2007), network design (Atamtürk and Zhang 2007), finance (Fabozzi et al. 2010, Bertsimas and Pachamanova 2008), industrial design (Den Hertog and Stehouwer 2002), and many other fields. For a broad overview of applications of RO, the reader is referred to Gabrel et al. (2014).

In the following, two sub-fields of RO are introduced that are relevant to this dissertation. In Section 1.2, the Adjustable Robust Optimization (ARO) is introduced - an extension of RO to multi-stage setting and where the aim is to make the later-stage decisions dependent on the revealed uncertainties from previous periods in a computationally tractable way. In Section 1.3, Distributionally Robust Optimization (DRO) is introduced where the uncertain parameter is the probability distribution of $\boldsymbol{z}$. Section 1.4 includes an overview of the research papers included in this thesis.

1.2 Adjustable Robust Optimization

It is a property of many real-life applications of optimization that the decisions are made over several time periods, between which more and more of the initially uncertain input of the problem becomes known. An illustrative two-stage RO problem can be formulated as:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}_1,\boldsymbol{x}_2} & t \\ \text{s.t.} & f_0(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{z}) \leq t, \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\ & f_i(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{z}) \leq 0, \qquad i=1,\ldots,m, \quad \forall \boldsymbol{z} \in \mathcal{Z}, \end{array} \tag{1.6}$$

where $\boldsymbol{x}_1$ is the decision vector implemented at time 1, before the uncertain parameter $\boldsymbol{z}$ is known, and $\boldsymbol{x}_2$ is the decision vector implemented at time 2, after $\boldsymbol{z}$ becomes known.

Problem (1.6) imposes rather restrictive limitations for the time 2 decisions $\boldsymbol{x}_2$, since they are supposed to be determined without any knowledge of $\boldsymbol{z}$ and thus, can turn out to be rather conservative/suboptimal for particular outcomes of $\boldsymbol{z}$. To mitigate this shortcoming, the idea of ARO is to formulate the second stage decisions as a function of $\boldsymbol{z}$, and to choose the best 'response function'. Mathematically, this is formulated as:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}_1,\boldsymbol{x}_2(\cdot)} & t \\ \text{s.t.} & f_0(\boldsymbol{x}_1,\boldsymbol{x}_2(\boldsymbol{z}),\boldsymbol{z}) \leq t, \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\ & f_i(\boldsymbol{x}_1,\boldsymbol{x}_2(\boldsymbol{z}),\boldsymbol{z}) \leq 0, \qquad i=1,\ldots,m, \quad \forall \boldsymbol{z} \in \mathcal{Z}. \end{array} \tag{1.7}$$

Now that $\boldsymbol{x}_2(\boldsymbol{z})$ can adjust to $\boldsymbol{z}$, the optimal value is expected to be better compared to (1.6). Unfortunately, problem (1.7) is $\mathcal{NP}$-hard in the general case, and the main task of ARO is to identify 'how close' one can get to this ideal solution.

Initially, ARO was developed to solve problems with continuous variables. Ben-Tal et al. (2004) introduced the concept of using affinely adjustable decision rules and show how to apply such rules to obtain (approximate) optimal solutions to multiperiod problems. Their approach has been later extended to other function classes by Chen et al. (2007), Chen and Zhang (2009), Ben-Tal et al. (2009) and Bertsimas et al. (2011).

Later, initial developments have been made allowing ARO to (approximately) solve problems involving adjustable integer variables, as in Bertsimas and Caramanis (2007), Vayanos et al. (2011), Bertsimas and Georghiou (2015), Bertsimas and Georghiou (2014), and Hanasusanto et al. (2015a).

The main challenges of ARO are, among others:

- How to provide scalable methods for adjustable integer decisions?

- How to reduce the computational burden related to the fact that even with relatively scalable methods, one needs to solve mixed-integer problems?
- What is the impact of including integer adjustability on the here-and-now decisions in important applications of optimization?

In Chapter 3 a novel technique is proposed to construct iteratively improving integer decisions in multi-stage problems, based on identifying the scenarios for the uncertain parameter $\boldsymbol{z}$, for which decisions should be different. In Chapter 5 we study multi-stage distributionally robust stochastic programming problems with integer recourse, for which we are able to provide their convex approximations with explicit performance bounds. Chapter 5 includes also applied techniques showing how the resulting problems can be solved in a computationally efficient way. In Chapter 6 we apply the integer adjustability approach of Chapter 3 to construct adjustable robust water flood protection strategies in the Netherlands. The numerical results clearly show that the inclusion of integer adjustability is of great practical importance already for the here-and-now decisions, which can be better when the future readjustment possibilities are taken into account.

1.3 Distributionally Robust Optimization

In certain applications, the uncertain parameter $\boldsymbol{z}$ is known to be stochastic and to require that a constraint holds for all possible outcomes of $\boldsymbol{z} \in \mathcal{Z}$ where $\mathcal{Z}$ is the support of $\boldsymbol{z}$, would be (i) too conservative (for example, when the random variable has $\mathbb{R}^n$ as its support), or (ii) impossible to reformulate to a computationally tractable form.

At the same time, some information is known about the probability distribution $\mathbb{P}$ of $\boldsymbol{z}$, which gives rise to an entire set $\mathcal{P}_{\boldsymbol{z}}$ of probability distributions $\mathbb{P}_{\boldsymbol{z}}$ satisfying this information, that is, a set of probability distributions all of which can be true under what is known. Then, one can consider constraints that use this probabilistic information. The two most common constraints are the worst-case expected feasibility constraints:

$$\mathbb{E}_{\mathbb{P}_{\boldsymbol{z}}} f(\boldsymbol{x}, \boldsymbol{z}) \leq 0, \quad \forall \mathbb{P}_{\boldsymbol{z}} \in \mathcal{P}_{\boldsymbol{z}}, \tag{1.8}$$

where some overload of notation is applied since, strictly speaking, if the uncertain parameter is $\mathbb{P}_{\boldsymbol{z}}$, then the constraint should be in the form $f(\boldsymbol{x}, \mathbb{P}_{\boldsymbol{z}})$ but in this way we stay consistent with the notation of the later chapters, and the worst-case chance-constraints:

$$\mathbb{P}_{\boldsymbol{z}}(f(\boldsymbol{x}, \boldsymbol{z}) > 0) \leq \epsilon, \quad \forall \mathbb{P}_{\boldsymbol{z}} \in \mathcal{P}_{\boldsymbol{z}}. \tag{1.9}$$

Note that (1.9) is a special case of (1.8) if the function $f(\boldsymbol{x}, \boldsymbol{z})$ in (1.8) is an indicator function, but it is often studied separately due to its difficulty and practical importance.

Research on constraints of types (1.8) and (1.9) goes under the name Distributionally Robust Optimization (DRO), introduced first in Delage and Ye (2010), who study under what conditions the constraints can be reformulated either exactly, or to their conservative approximations - systems of constraints such that $\boldsymbol{x}$ satisfying the conservative approximation also satisfies (1.8) or (1.9).

Information about the distribution of $\boldsymbol{z}$ can be formulated in terms of:

- support: discrete, continuous, bounded, unbounded, etc;
- moments: mean, covariance, higher order moments;
- structural information: symmetry, unimodality, independence of components;
- distance from a known distribution, as measured by means of test statistics or other probability metrics.

For a broad overview of types of partial information considered in the literature the reader is referred to Hanasusanto et al. (2015b).

It should be added that both (1.8) and (1.9) have been studied already within the framework of Stochastic Optimization (SP), see Birge and Louveaux (1997). In the SP setting, however, the stress has been much more on techniques of bounding the expectations from above or using sampling techniques ensuring that the constraints are satisfied with high accuracy. In RO, on the other hand, most of the research concentrates on using duality to obtain tractable reformulations. In this sense, the two approaches can be seen as complementary (in very specific cases the difference can be purely interpretational) and in recent years we can observe their convergence.

DRO is an attractive paradigm for its capacity to minimize/maximize expectations of the objective functions rather than the worst-case objective function values, as in problem (1.2). This is because of the fact that optimization is often applied in fields in which the processes are repeatable, such as inventory management, production, network flow optimization, where cost/profit in the long run converges to the average.

Up to now, the research community has been able to provide many tractable reformulations mostly for the worst-case expected feasibility constraints (1.8). Delage and Ye (2010) show that in many practicable situations, worst-case expected feasibility constraints can be tractable under mean-second order moment uncertainty. Wiesemann et al. (2014) proposed a unifying framework for deriving robust counterparts of problems where the uncertain parameter is the probability distribution, using duality for moment problems. For distance-based uncertainty, Ben-Tal et al. (2013)

provide such reformulations for discrete probability distributions with uncertainty sets defined using the so-called φ-divergence goodness-of-fit statistics.

Chance constraints, on the other hand, are more difficult as the uncertainty quantification problem, that is, the task of determining the worst-case violation probability of a single constraint becomes $\mathcal{NP}$-hard in even very simple settings (Nemirovski and Shapiro 2006). In some of the earliest papers El Ghaoui et al. (2003) show how worst-case constraints on the Value-at-Risk of linear expressions (equivalent to chance-constraints) can be obtained under uncertainty about the mean and covariance matrix of $\boldsymbol{z}$. These safe approximations are extended, and several are proposed in the monograph of Ben-Tal et al. (2009). Under second-order moment information, Zymler et al. (2011) construct safe approximations of *joint chance constraints*, where a number of different constraints is to hold jointly with probability at least $1-\epsilon$. A recent work aiming at providing a more general framework of what is tractable and what is intractable for joint chance constraints is Hanasusanto et al. (2015b).

Among papers considering uncertainty in terms of distance from a known distribution, Jiang and Guan (2015) show exact reformulations of individual constraints under the Kullback-Leibler divergence. Esfahani and Kuhn (2014) show how distributionally robust chance constraints can be approximated under the Wasserstein distance in a data-driven setting.

The key challenges facing the DRO field are, among other things:

- How to construct tractable uncertainty sets for general functions in a data-driven setting, both for discrete and continuous probability distributions?
- How to obtain computationally tractable decisions problems that explicitly take into account the distributional ambiguity?
- What is the value of the additional distributional information that one could obtain in order to improve the solutions?

In Chapter 2 we show how to construct statistics-based uncertainty sets for discrete probability measures and how to derive computationally tractable robust counterparts of constraints on risk measures of decision-dependent random variables. In Chapter 4 we consider continuous random variables under mean - mean absolute deviation distributional information, which can be easily estimated from the data. We show how under this information, computationally tractable robust counterparts of constraints (1.8) and (1.9) can be derived. This information is also used in Chapter 5 where we construct computationally tractable robust counterparts of multi-stage mixed-integer stochastic programming problems. This approach allows us to explicitly evaluate the value of distributional information in the form of tight bounds on

the worst- and best-case expectations of the objective value under the given distributional information.

1.4 Overview of research papers

The rest of this thesis consists of five self-contained chapters. Here, the contributions of each chapter are given.

In Chapter 2 an important class of distributionally robust constraints is identified for which tractable reformulations can be found using Fenchel duality: constraints on risk measures of decision dependent discrete random variables under statistics-based uncertainty sets for the probability vector. In particular, we show how to construct such counterparts for certain nonlinear risk measures. Our integrated framework covers and extends beyond the numerous existing results in this field.

In Chapter 3 a novel approach is developed for constructing iteratively refinable integer decisions for multi-stage problems. The idea is to split the uncertainty set into subsets, each of which is assigned to a different decision. This technique is based on the observation that after a RO problem is solved, different constraints are 'made active' by different realizations of the uncertain parameter. Theoretical results are derived that confirm this observation and give guidance on how to split the uncertainty set. Additionally, this technique can be combined with existing linear decision rules, allowing to differentiate the affine rules in different parts of the uncertainty set.

In Chapter 4 we consider distributionally robust constraints of both types under mean and mean-absolute deviation information about the stochastically independent components of a random variable. We show that for such a setting, we can leverage old results from SP on the expectations of convex functions to obtain new, closed-form equivalent robust formulations of worst-case expected feasibility constraints or safe approximations of chance constraints. In addition, we identify an important class of problems where we can relax the assumption on independence of the components of the random vector. The approach can be used, among others, to enhance the average-case performance of ARO solutions in situations where there exist multiple optimal solutions or to solve problems involving implementation error.

In Chapter 5 we combine the distributionally robust and integer adjustable fields, studying multi-stage stochastic programming problems under distributional uncertainty, under the same distributional information as in Chapter 4. We propose convex approximations of the problems with integer later-stage variables for which explicit performance bounds are derived, and show how numerous techniques can be used to reduce the computational effort related to the distributional uncertainty. This approach allows for explicit evaluation of the value of distributional information in

the form of tight bounds on the worst- and best-case expectations of the objective value under the given distributional information.

Chapter 6 is an application of integer adjustable RO to the problem of optimizing flood protection measures in the Rhine-Meuse Estuary area in the Netherlands. We show how explicit inclusion of uncertainty and adjustability can lead to finding better here-and-now investment decisions and an overall reduction of the investment costs.

1.5 Disclosure

This thesis is based on the following five research papers:

Chapter 2 K. Postek, D. den Hertog and B. Melenberg. Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review*, 2016.

Chapter 3 K. Postek and D. den Hertog. Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing*, 2016.

Chapter 4 K. Postek, A. Ben-Tal, D. den Hertog and B. Melenberg. Robust counterparts of ambiguous stochastic constraints under mean and dispersion information. 2015. In second review round for publication at *Operations Research*.

Chapter 5 K. Postek, W. Romeijnders, D. den Hertog, and M. van der Vlerk. Efficient methods for several classes of ambiguous stochastic programming problems under mean-MAD information. 2016. CentER Discussion Paper No. 2016-039. Submitted.

Chapter 6 K. Postek, D. den Hertog, J. Kind, Ch. Pustjens. Adjustable robust strategies for flood protection. 2016. CentER Discussion Paper No. 2016-038. Submitted.

Each chapter contains ideas and contributions from all its respective authors. In Chapters 2, 3, 4, and 6 all the sections are written by me and all the experiments are done by me. In Chapter 5 Sections 5.1, 5.2, 5.5.3, and 5.6 are written by me and the corresponding experiments are done by me.

Bibliography

A. Atamtürk and M. Zhang. Two-stage robust network row and design under demand uncertainty. *Operations Research*, 55:662–673, 2007.

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

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

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

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

A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.

D. Bertsimas and C. Caramanis. Adaptability via sampling. *46th IEEE Conference on Decision and Control*, pages 4717–4722, 2007.

D. Bertsimas and A. Georghiou. Binary decision rules for multistage adaptive mixed-integer optimization. http://www.optimization-online.org/DB_FILE/2014/08/4510.pdf, 2014.

D. Bertsimas and A. Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *IEEE Transactions on Power Systems*, 63(3):610–627, 2015.

D. Bertsimas and D. Pachamanova. Robust multiperiod portfolio management in the presence of transaction costs. *Computers & Operations Research*, 35(1):3–17, 2008.

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

D. Bertsimas, D. Pachamanova, and M. Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.

D. Bertsimas, D. A. Iancu, and P. A. Parrilo. A hierarchy of near-optimal policies for

multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12): 2809–2824, 2011.

D. Bienstock. Histogram models for robust portfolio optimization. *Journal of Computational Finance*, 11(1), 2007.

J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2009.

X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

D. Den Hertog and P. Stehouwer. Optimizing color picture tubes by high-cost nonlinear programming. *European Journal of Operational Research*, 140(2):197–211, 2002.

L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.

L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. http://arxiv.org/abs/1505.05116, 2014.

F. J. Fabozzi, D. Huang, and G. Zhou. Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176:191–220, 2010.

V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483, 2014.

D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage robust binary programming. *Operations Research*, 64(3):877–891, 2015a.

G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62, 2015b.

R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Advance online publication in Mathematical Programming*, 2015.

R. Laraki and J. B. Lasserre. Semidefinite programming for N-player games. http://arxiv.org/pdf/0810.3150.pdf, 2008.

J. B. Lasserre. *Moments, positive polynomials and their applications.* Imperial College Press, 2009.

J. B. Lasserre. *Introduction to Polynomial and Semi-Algebraic Optimization.* Cambridge University Press, 2015.

A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

F. Ordonez and J. Zhao. Robust capacity expansion of network flows. *Networks*, 50: 136–145, 2007.

A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

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

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1):167–198, 2011.

CHAPTER 2

Computationally tractable counterparts of distributionally robust constraints on risk measures

2.1 Introduction

Robust Optimization (RO, see Ben-Tal et al. [8]) has become one of the main approaches to optimization under uncertainty. Since its introduction it has already found numerous applications. Its paradigm lies in assuming that some or all of the optimization problem's parameters are uncertain. For these parameters an uncertainty set is specified. This uncertainty set includes parameter values for which the solution should be feasible. The robust optimization problem is then solved in such a way that the solution is best possible for the worst-case parameter values from the uncertainty set. For an introduction and overview of techniques used in RO, we refer the reader to the work by Bertsimas and Brown [12] and references therein. A more recent survey of the developments and applications of RO is Gabrel et al. [25].

A particular application field for RO is keeping risk measures of decision-dependent random variables below pre-specified limits. Such constraints typically appear in finance, engineering, and economics. Often, the computation of the value of a risk measure requires knowledge of the underlying probability distribution, which is usually estimated. Such an estimate is typically based on a number of past observations. Due to sampling error, this estimate approximates the true distribution only with a limited accuracy. A confidence set around the estimate gives rise to a natural uncertainty set of admissible probability distributions (at a given confidence level). The key difficulty lies in reformulating the problem's constraints in a way that allows for the application of efficient optimization algorithms. Such a reformulation is referred to as a tractable robust counterpart of the constraint (see Ben-Tal et al. [8]). Many authors study this type of distributional uncertainty and the corresponding tractable robust counterparts. In Section 2.2 we present a review of existing results. Typically, they focus on a specific combination of risk measure and uncertainty set.

In this paper, apart from providing an overview of the results in the literature, we propose a unified approach to derive computationally tractable robust counterparts of this kind of constraints. Our approach allows us to deal with many more risk measure-uncertainty set combinations than have been considered up to now. The unifying approach of our paper consists of the following three parts.

First, using Fenchel duality and results of Ben-Tal et al. [10] we show that the derivation of the tractable robust counterpart can be separated in terms of the components corresponding to the risk measure and the uncertainty set. Therefore, we derive two types of building blocks: one for the risk measures and another for the uncertainty sets. The resulting blocks may be combined arbitrarily according to the problem at hand. This provides the decision maker with a unified structure to reformulate this type of constraints, allowing to cover many more risk measure-uncertainty set combinations than is captured up to now in the literature. The first building block includes the following risk measures: negative mean return, Optimized Certainty Equivalent (with Conditional Value-at-Risk as a special case), Certainty Equivalent, Shortfall Risk, lower partial moments, mean absolute deviation from the median, standard deviation/variance less mean, Sharpe ratio, and the Entropic Value-at-Risk. The second building block encompasses uncertainty sets built using φ-divergences (with the Pearson (χ^2) and likelihood ratio (G) tests as special cases), Kolmogorov-Smirnov test, Wasserstein (Kantorovich) distance, Anderson-Darling, Cramer-von Mises, Watson, and Kuiper tests.

Secondly, we address a common feature rendering many optimization constraints computationally difficult, namely, nonlinearity of some of the risk measures in the underlying probability distribution. In our setting, this is the case for the variance, the standard deviation, the Optimized Certainty Equivalent, and the mean absolute deviation from the median. To make the use of RO methodology possible, we use different, equivalent reformulations of such risk measures as infima over relevant function sets, whose elements are linear in the probabilities. A minmax result from convex analysis ensures that this operation results in an exact reformulation.

Thirdly, we provide the complexity status (linear, convex quadratic, second-order conic, convex) of the robust counterparts. This is summarized in Table 2.1, together with a summary of the results captured in the literature up to now. As illustrated, our methodology allows for obtaining a tractable robust counterpart for most of the risk measure-uncertainty set combinations, extending the results in the field.

For several types of risk measures, including the Value-at-Risk, the mean absolute deviation from the mean, and general-form distortion, coherent and spectral risk measures, we could not reformulate the relevant constraints into a tractable form within our framework. Section 2.5 contains a brief discussion of reasons why our approach is not applicable to such cases.

We remark that another type of distributional uncertainty in RO problems is uncertainty in the moments of some key random variables, as studied, e.g., by El Ghaoui et al. [20] and Delage and Ye [18]. However, in our paper we focus on uncertainty in the discrete probabilities of these random variables.

The composition of the remainder of the paper is as follows. Section 2.2 provides a survey of the results obtained in the literature so far. Section 2.3 introduces the definitions and the main tool for deriving the computationally tractable robust counterparts. Section 2.4 lists the risk measures and uncertainty sets for the probability distribution that we investigate. Sections 2.5 and 2.6 include the results on the building blocks of the robust counterparts of the constraints on the risk measures. In Section 2.7, we provide examples of combining the blocks, including a numerical study. Section 2.8 concludes and lists the potential directions for future research.

2.2 Literature review

In this section we present the results of existing research on deriving tractable forms of distributionally robust constraints on risk measures, grouped according to the type of uncertainty sets, which is in line with Table 2.1. To provide the reader with a better understanding of the literature review, we first present as example a constraint on the standard deviation.

In our setting, it is assumed that the decisions form a vector w and the decision-dependent random variable $X(w)$, representing the decision maker's reward, has possible outcomes $X_1(w), \ldots, X_N(w)$ with probabilities $p_1, \ldots, p_N$, respectively. The probability vector $p = [p_1, \ldots, p_N]^T$ is known to belong to a confidence region $\mathcal{P}$. For such a random variable, we want its standard deviation not to be greater than β for all possible probability distributions in $\mathcal{P}$. The constraint is then:

$$\sqrt{\sum_{n=1}^{N} p_n \left(X_n(w) - \sum_{n'=1}^{N} p_{n'} X_{n'}(w) \right)^2} \leq \beta, \quad \forall p \in \mathcal{P}. \tag{2.1}$$

The key question is whether (2.1) can be reformulated into an equivalent form that does not have the semi-infinite 'for all' form. Such combinations of uncertain discrete probabilities and risk measures have already been investigated by several authors.

Calafiore [16] studies a portfolio optimization problem with risk defined as the variance less the mean and mean absolute deviation from the mean, under distributional uncertainty defined with the use of Kullback-Leibler divergence. He notices that the step of finding the worst-case probability distribution for a given portfolio can be conducted efficiently. Combining it with the generation of cutting planes for the general robust optimization problem, he proposes an algorithm that finds the optimal portfolios in polynomial time.

Table 2.1 – Results on complexity of a tractable counterpart for risk measures and uncertainty sets. The symbol • means that a tractable robust counterpart has been formulated in the literature and the symbol ∘ means that only a partial solution was found in the literature, e.g., an efficient method of evaluating the worst-case values, or that the tractable robust counterpart can be obtained from authors' results with some additional work. The complexity symbols are: LP - linear constraints, QP - convex quadratic, SOCP - second-order conic, CP - convex. The symbol ∗ means that the right-hand side in a constraint (β in constraint (2.2)) must be a fixed number for the counterpart to be a system of convex constraints. The results are constructed assuming that the decision-dependent random variable $X(w)$ is linear in the decision vector w (see Section 2.4).

Uncertainty set → **Risk measure ↓**	φ-divergences	Pearson	Likelihood ratio	Kolmogorov-Smirnov	Wasserstein (Kantorovich)	Anderson-Darling	Cramer-von Mises	Watson	Kuiper
Negative mean return	• CP [10],[9],[15]	• SOCP [10][9][33][50]	• CP [10][9][29][49]	∘ LP [50]	• LP [40][50][51]	• CP [10]	∘ SOCP [50]	∘ SOCP [50]	∘ LP [50]
Standard deviation less the mean	CP	SOCP	CP	SOCP	• SOCP [51]	CP	SOCP	SOCP	SOCP
Standard deviation	CP	SOCP	CP	SOCP	• SOCP [51]	CP	SOCP	SOCP	SOCP
Variance less the mean	CP	SOCP	∘ CP [16]	QP	QP	CP	SOCP	SOCP	QP
Variance	∘ CP [10]	∘ SOCP [10]	∘ CP [10][16]	QP	QP	∘ CP [10]	SOCP	SOCP	QP
Mean absolute deviation from the median	CP	SOCP	CP	LP	• LP [51]	CP	SOCP	SOCP	LP
Sharpe ratio	CP∗	SOCP∗	CP∗	SOCP∗	SOCP∗	CP∗	SOCP∗	SOCP ∗	SOCP ∗
Lower partial moment $\alpha = 1$	∘ CP [10],[9]	∘ SOCP [10][9][33][50]	∘ CP [10][9][29][49]	∘ LP [50]	∘ LP [50]	∘ CP [10]	∘ SOCP [50]	∘ SOCP [50]	∘ LP [50]
Lower partial moment $\alpha = 2$	∘ CP [10],[9]	∘ SOCP [10][9][33]	∘ CP [10][9][29][49]	QP	QP	∘ CP [10]	SOCP	SOCP	QP
Certainty Equivalent	CP∗	CP∗	∘ CP ∗	CP∗	CP∗	CP∗	CP∗	CP∗	CP∗
Shortfall risk	∘ CP [10],[9]	∘ CP [10][9][33]	∘ CP [10][9][29][49]	CP	CP	∘ CP [10]	CP	CP	CP
Optimized Certainty Equivalent	CP	CP	CP	CP	CP	CP	CP	CP	CP
Value-at-Risk	∘ [30]	∘ [30]	∘ [29][30][32]						
Conditional Value-at-Risk	CP	∘ SOCP [13][50]	• CP [13][29]	∘ LP [13][50]	• LP [40][50][51]	∘ CP [13]	∘ SOCP [13][50]	∘ SOCP [13][50]	∘ LP [13][50]
Entropic Value-at-Risk	CP	CP	CP	CP	CP	CP	CP	CP	CP
Mean absolute deviation from the mean			∘ [16]						
Distortion risk measures					• [40][51]				
Coherent risk measures									
Spectral risk measures									

Jiang and Guan [32] consider ambiguous chance constraints under the Kullback-Leibler divergence, reducing the robust chance-constrained problem to a problem under the nominal probability measure, with modified violation probability. In our setting, these results have applications to constraints on the Value-at-Risk (VaR), whose equivalence to chance constraints has been noted already by Nemirovski and Shapiro [39]. Hu and Hong [29] also consider optimization problems with convex expectation constraints under distributional uncertainty defined with the Kullback-Leibler divergence. They provide closed-form distributionally robust counterparts of constraints on expectations of general convex performance measures. Their results apply, for example, to Conditional Value-at-Risk (CVaR) and, as an approximation, to VaR. Results of Hu and Hong [29] are partly generalized in Hu et al. [30], who consider chance-constrained problems with distributional uncertainty sets defined by general φ-divergence functions. They also show that the robust constraints are equivalent to nominal constraints with modified confidence levels. Another work, that does not concentrate on risk measures as such, including derivations of tractable robust counterparts under distributional uncertainty, is Klabjan et al. [33]. In this work, a lot-sizing problem with uncertainty defined with the χ^2-test statistic is solved. After an appropriate reformulation, the problem solved is a SOCP problem.

Wang et al. [49] derive tractable counterparts of constraints involving linear functions of the probability vector, with uncertainty defined by the likelihood ratio test. They also provide various interpretations of the obtained result, for example, from the Bayesian perspective. Ben-Tal et al. [9] study constraints on general convex functions under distributional uncertainty defined by φ-divergence functions, generalizing the result of Wang et al. [49]. They derive tractable robust counterparts, showing that for several of the divergences the resulting constraint allows for a self-concordant barrier function. Their results, as a specific case, apply to such risk measures as the negative mean return. Methods for obtaining worst-case probability distributions under φ-divergence uncertainty are given also in Breuer and Csiszár [15].

Wozabal [51] studies portfolio optimization with risk measures such as expectation, standard deviation less the mean, mean absolute deviation from the median, CVaR, distortion risk measure, Wang transform, proportional hazards transform, and the Gini measure, under distributional uncertainty defined with the Wasserstein distance. Using the so-called subdifferential representation of risk measures, he derives closed-form worst-case values of risk measures. For the first three risk measures the resulting worst-case expressions are linear, convex quadratic, or piecewise linear in the decision variables. Pichler [40] focuses on the worst-case values of general spectral and distortion risk measures under distributional uncertainty defined with the Wasserstein distance. He provides expressions for the so-called transport maps that define the worst-case probability distributions for given values of the decision variables. However, in a general case these formulas cannot be implemented easily because of their

nonlinear (nonconvex) forms.

There are also several works studying risk measures or uncertainty sets different from the ones we consider. Zhu and Fukushima [52] analyze the CVaR under box and ellipsoidal uncertainty sets for discrete probabilities. Using the min-formulation for CVaR from Rockafellar and Uryasev [43] and minmax results from convex analysis, they show that the problem of minimizing the worst-case CVaR in such a case can be formulated as LP (for box uncertainty) or SOCP (for ellipsoidal uncertainty). Fertis et al. [23] study the CVaR under an uncertainty set with a two-stage structure. In this structure, the uncertainty set is defined as a ball with an arbitrary norm around a reference probability distribution. This reference probability distribution is allowed to be a convex combination of a finite number of known probability distributions. In this way, the authors generalize the results of Zhu and Fukushima [52] to the continuous case, showing that a constraint on the CVaR can be reformulated tractably to a system of constraints involving dual norms.

Huang et al. [31] propose a framework replacing the standard CVaR by a less conservative measure, namely the Relative Robust CVaR, and show that under a multiple-expert uncertainty set the resulting optimization problem can be reformulated either as LP or as SOCP. Mosler and Bazovkin [37] construct a geometrically-based method for solving robust linear programs with a single distortion risk measure under polytopial uncertainty sets. It is not known yet whether their results can be extended to the statistically-based uncertainty sets for probabilities.

Wiesemann et al. [50] propose a general framework of distributionally robust convex optimization. They require that the function to be constrained is bilinear in the decision variables and the random vector, imposing an uncertainty set that must possess a conic representation, with some regularity conditions. The reformulation they provide applies to multiple risk measures and conic-representable uncertainty sets for the probabilities, see Table 2.1 for an overview.

Ben-Tal et al. [10] provide the mathematical framework used in our paper. They study general nonlinear robust constraints reformulated using Fenchel duality. Results of their paper allow to obtain tractable constraints for the variance with distributional uncertainty defined by φ-divergences and the Anderson-Darling goodness of fit tests. In both cases the resulting system of constraints is convex, and for some of the φ-divergence functions it is second-order conic. In our paper we extend their framework to other types of uncertainty sets and risk measures.

Bertsimas et al. [13] show how in a data-driven setting one can construct uncertainty sets based on statistical tests such as Kolmogorov-Smirnov, χ^2, Anderson-Darling, Watson and likelihood ratio, used to obtain conservative bounds on the VaR via CVaR. They utilize a cutting plane algorithm with an efficient method of evaluating the worst-case values of the decision-dependent random variables.

Finally, two works, Natarajan et al. [38] and Bertsimas and Brown [11] provide a more general insight into the relation between robust optimization and risk measurement. They show that there is a one-to-one relationship between coherent risk measures (see Artzner et al. [4] for an introduction) and uncertainty sets for general uncertain parameters in the case of constraints that are linear in this uncertainty.

2.3 Preliminaries

As already introduced in Section 2.2, we study constraints on risk measures of decision-dependent random variables, where $w \in \mathbb{R}^M$ is the decision vector. The random variable $X(w)$, representing the decision maker's reward, and whose risk is measured, takes a value $X_n(w)$ with probability p_n for each $n \in \mathcal{N} = \{1, \ldots, N\}$. The uncertain parameter is the probability vector $p = [p_1, ..., p_N]^T \in \mathbb{R}^N_+$, representing the discrete distribution of $X(w)$. The reference probability vector, around which the uncertainty set for p may be specified, is denoted by $q \in \mathbb{R}^N_+$.

Let the risk measure of the random variable $X(w)$ under the probability distribution represented by the vector p be given by $\overline{F}(p, X(w))$, with $\overline{F} : \mathbb{R}^N_+ \times \mathbb{R}^N \to \mathbb{R}$. The robust constraint on the risk measure that we shall reformulate to a tractable form is given by:

$$F(p, w) = \overline{F}(p, X(w)) \leq \beta, \quad \forall p \in \mathcal{P}, \tag{2.2}$$

where $F : \mathbb{R}^N_+ \times \mathbb{R}^M \to \mathbb{R}$ and $\mathcal{P}$ is the uncertainty set for the probabilities defined as:

$$\mathcal{P} = \{p : \quad p = Ap', \quad p' \in \mathcal{U}\}, \tag{2.3}$$

where the set $\mathcal{U} \subseteq \mathbb{R}^L_+$ is a nonempty, compact convex set, and $A \in \mathbb{R}^{N \times L}$ such that $\mathcal{P} \subseteq \mathbb{R}^N_+$. The formulation of the set $\mathcal{P}$ using the matrix A is general and encompasses cases where the set $\mathcal{U}$ has a dimension different from N.

Example 2.1 *If the risk measure of the random variable $X(w)$ is the standard deviation and the uncertainty set is defined using a φ-divergence function around the reference probability vector q (see Table 2.3), then the constraint is:*

$$F(p, w) = \overline{F}(p, X(w)) = \sqrt{\sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2} \leq \beta, \quad \forall p \in \mathcal{P},$$

with $A = I$ and

$$\mathcal{P} = \mathcal{U} = \left\{ p \in \mathbb{R}^N_+ : \quad \sum_{n \in \mathcal{N}} p_n = 1, \quad \sum_{n \in \mathcal{N}} q_n \varphi\left(\frac{p_n}{q_n}\right) \leq \rho \right\}.$$

The robust constraint of Example 2.1 is reformulated to a computationally tractable form in Section 2.7.1.

We now introduce the key theorem used in this paper to construct a unified framework for tackling constraints involving a risk measure and an uncertainty set for probabilities. First, we give the definition of the conjugate functions and the support function, adapted to our context. The *concave conjugate* $f_*(\cdot)$ of a function $f : \mathbb{R}_+^N \to \mathbb{R}$ is defined as a function $f_* : \mathbb{R}^N \to \mathbb{R} \cup \{-\infty\}$:

$$f_*(v) = \inf_{p \in \mathbb{R}_+^N} \left\{ v^T p - f(p) \right\}. \tag{2.4}$$

The *convex conjugate* $g^*(\cdot)$ of a function $g : \mathbb{R}_+^N \to \mathbb{R}$ is defined as a function $g^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$:

$$g^*(v) = \sup_{p \in \mathbb{R}_+^N} \left\{ v^T p - g(p) \right\}. \tag{2.5}$$

Remark 2.1 *In the above definitions, the domains of $f(\cdot)$ and $g(\cdot)$ are given by $\mathbb{R}_+^N$, and the corresponding conjugates are defined as an infimum/supremum over $\mathbb{R}_+^N$, instead of $\mathbb{R}^N$. Our approach can easily be adapted in a way that fits the standard definitions by setting the values of $f(\cdot)$ and $g(\cdot)$ equal to $-\infty$ and $+\infty$ outside $\mathbb{R}_+^N$, respectively, so that $\mathbb{R}_+^N$ is the effective domain.*

The indicator function $\delta(\cdot|\mathcal{U})$ of a nonempty set $\mathcal{U} \subseteq \mathbb{R}_+^L$ is defined as

$$\delta(p'|\mathcal{U}) = \begin{cases} 0 & \text{if } p' \in \mathcal{U} \\ +\infty & \text{otherwise,} \end{cases}$$

and the *support function* $\delta^*(\cdot|\mathcal{U}) : \mathbb{R}^L \to \mathbb{R} \cup \{\infty\}$ of a nonempty set $\mathcal{U} \subseteq \mathbb{R}_+^L$ is defined as a convex conjugate of the indicator function:

$$\delta^*(v|\mathcal{U}) = \sup_{p' \in \mathbb{R}_+^L} \left(v^T p' - \delta(p'|\mathcal{U}) \right) = \sup_{p' \in \mathcal{U}} v^T p'. \tag{2.6}$$

The following theorem, adapted from Ben-Tal et al. [10], is the main tool for deriving the tractable robust counterparts in this paper.

Theorem 2.1 *Let $f : \mathbb{R}_+^N \times \mathbb{R}^M \to \mathbb{R}$ be a function such that $f(\cdot, w)$ is closed and concave for each $w \in \mathbb{R}^M$. Consider a constraint of the form:*

$$f(p, w) \leq \beta, \quad \forall p \in \mathcal{P}, \tag{2.7}$$

where $\mathcal{P}$ is defined by (2.3) and where it holds that:

$$ri(\mathcal{P}) \cap \mathbb{R}_{++}^N \neq \emptyset. \tag{2.8}$$

Then (2.7) holds for a given w if and only if:

$$\exists v \in \mathbb{R}^N : \quad \delta^* \left(A^T v \middle| \mathcal{U} \right) - f_*(v, w) \leq \beta, \tag{2.9}$$

where $\delta^*(\cdot|\mathcal{U})$ is the support function of the set $\mathcal{U}$ and $f_*(\cdot, w)$ is the concave conjugate of $f(\cdot, w)$ with respect to its first argument.

Proof. The proof relies on the Fenchel duality theorem, included as Theorem 2.2 in Appendix 2.A. Constraint (2.7) is equivalent to:

$$G(w) = \sup_{p \in \mathcal{P}} \{ f(p, w) - \delta \left(p \,|\mathcal{P} \right) \} \leq \beta.$$

We have that

$$\begin{aligned} G(w) &= \sup_{p \in \mathcal{P}} \{ f(p, w) - \delta \left(p \,|\mathcal{P} \right) \} \\ &= \inf_v \{ \delta^* \left(v | \,\mathcal{P} \right) - f_*(v, w) \} \\ &= \inf_v \left\{ \delta^* \left(A^T v \middle| \mathcal{U} \right) - f_*(v, w) \right\}, \end{aligned}$$

where the second equality follows from Fenchel duality. Moreover, due to condition (2.8) the infimum is attained, see Appendix 2.A. Because of this, the constraint

$$\inf_v \left\{ \delta^* \left(A^T v \middle| \mathcal{U} \right) - f_*(v, w) \right\} \leq \beta$$

is equivalent to (2.9), obtained by removing the inf term. ■

We argue in Section 2.4.2 that for the uncertainty sets considered in this paper condition (2.8) holds under mild conditions on the vector q. For cases where $f(\cdot, w)$ is not closed or concave, (2.9) is a conservative approximation of (2.7), which follows from weak Fenchel duality.

Theorem 2.1 allows for a separation of the derivation of two components: (i) the support function of the uncertainty set $\mathcal{U}$ at the point $A^T v$, and (ii) the concave conjugate of $f(\cdot, w)$. Moreover, it is not necessary to have closed-form formulations of the two components. For example, if one can express the support function as an infimum of some convex function over a set of parameter values, then the inf symbol can be removed after inserting such a formulation into (2.9), due to its position on the left hand side of the constraint. As we shall see in further sections, we often make use of this property.

If the function $F(\cdot, w)$ in (2.2) satisfies the concavity assumption with respect to p and we can obtain its conjugate directly from (2.5), then we can use Theorem 2.1 to reformulate the semi-infinite constraint (2.2) as a finite constraint of the type (2.9), setting $f(p, w) = F(p, w)$.

On the other hand, if the concavity assumption is not satisfied or the standard form of $F(\cdot, w)$ is too difficult to obtain a tractable conjugate, then we try to choose another function $f(\cdot, \cdot)$ such that (2.2) and (2.7) are equivalent, and Theorem 2.1 can be used.

Remark 2.2 *The form of the left-hand side in (2.9) is particularly useful when the conjugate and support functions are not available as closed-form expressions, but instead, are formulated as supremums and infimums, respectively:*

$$f_*(v, w) = \sup_{\lambda \in \Lambda(v,w)} \overline{f}(v, w, \lambda), \quad \delta^*(A^T v | \mathcal{U}) = \inf_{\theta \in \Theta(v)} \overline{g}(v, \theta),$$

where $\Lambda(v, w) \subseteq \mathbb{R}^{n_\Lambda}$ *and* $\Theta(v) \subseteq \mathbb{R}^{n_\Theta}$*. Inserting these formulations into (2.9) yields:*

$$\inf_{\theta \in \Theta(v)} \overline{g}(v, \theta) - \sup_{\lambda \in \Lambda(v,w)} \overline{f}(v, w, \lambda) \leq \beta.$$

Under the condition that the infimum of the left-hand side is attained, satisfied in examples considered in this paper, we obtain an equivalent formulation by removing the inf *and* sup *symbols, including the relevant constraints as:*

$$\exists (\lambda, \theta) \in \mathbb{R}^{n_\Lambda} \times \mathbb{R}^{n_\Theta} : \quad \begin{cases} \overline{g}(v, \theta) - \overline{f}(v, w, \lambda) \leq \beta \\ \lambda \in \Lambda(v, w) \\ \theta \in \Theta(v). \end{cases}$$

The next section gives the potential choices for the risk measures and the uncertainty set $\mathcal{P}$.

Additional notation

We distinguish the vectors by using the superscripts and the components of a vector using subscripts. For example, v_k^i denotes the k-th component of the vector v^i. Also, by the symbol $v_{s:t}$ we denote the subvector of v consisting of the components indexed s through t. Throughout the paper, 1 denotes a vector of ones, consistent in dimensionality with the equation at hand, 1^k is a vector with ones on its first k positions and zeros elsewhere , 1^{-k} is defined as the vector $1 - 1^k$, and e^k denotes a vector of zeros except a single 1 as the k-th component.

For formulas such as $f(p) = \sup_{h_i(x) \leq 0,\ i \in \mathcal{I}} g(x, p)$, we make use of the following layout:

$$\begin{aligned} f(p) = \sup \quad & g(x, p) \\ \text{s.t.} \quad & h_i(x) \leq 0,\ i \in \mathcal{I}, \end{aligned}$$

in situations where the terms under the sup/inf symbol would make the formulation difficult to read.

2.4 Risk measures and uncertainty sets

2.4.1 Risk measures

Risk measures, as concise quantifiers of the riskiness of random variables, can be designed for a great variety of purposes. The most important fields of their application are finance and economics (see, e.g., Schied and Föllmer [46] and Dowd [19]), linear regression in statistics (see, e.g., Rockafellar et al. [44]), supply chain management (see, e.g., Ahmed et al. [3]), engineering (see, e.g., Rockafellar and Royset [42]), and medicine (see, e.g., Calafiore [17]).

The choice of a risk measure depends on the risk characteristics one is interested in and is therefore likely to be application-specific. Also, each risk measure has its implications in terms of the tractable robust counterpart's complexity, as indicated in Table 2.1. In Table 2.2 we list a wide range of examples, exhausting a large share of practical applications. In their formulations, we follow the convention that the random variable $X(w)$ represents the reward and 'the smaller the risk measure, the better' (as a requirement for risk measures, known as the monotonicity axiom, see Artzner et al. [4]):

$$(\forall n \in \mathcal{N}: \quad X_n(w_1) \geq X_n(w_2)) \quad \Rightarrow \quad F(p, w_1) \leq F(p, w_2).$$

As an example, the first risk measure in Table 2.2 is the negative mean return, instead of its positive counterpart. In the remainder of this section we briefly introduce the included risk measures according to the aspect of risk they measure.

Some of the risk measures in Table 2.2 quantify the dispersion of the random variable $X(w)$ around a given 'central' level, such as the mean or the median. These risk measures include the variance, the standard deviation, and the mean absolute deviation from the median. The difference between, for example, the standard deviation and the variance lies in contributions made by deviations of different magnitude.

Variance or standard deviation minus the mean multiplied by a constant are popular in finance as they represent the classical 'return minus risk' performance measure, see Markowitz [36]. The Sharpe ratio (see Sharpe [47]), defined as the proportion of the mean of the random variable to its standard deviation is also popular in the financial context as a measure of the assets' riskiness. Typically, the mean used then is the mean excess return of an asset above the return of a riskfree asset. The combination of the standard deviation and the mean can also be used in engineering to ensure that the value of some random variable is not greater than a prescribed level by at least 'some number of standard deviations'.

Lower partial moments are useful when a one-sided deviation of the random variable around a specified level is important, as can be the case with losses in a financial setting. Lower partial moments with $\alpha = 1$ and $\alpha = 2$ differ in the contribution

made by deviations of different magnitude. The name 'lower partial moment' is just a convention and one can analyze similarly the upper partial moments of the variables.

To study not only the deviation around some given level, but the overall riskiness of a variable $X(w)$ in an economic context, risk measures involving the agent's utility function $u(\cdot)$ are used as well. These are (1) the Certainty Equivalent - the negative of the 'sure amount for which a decision maker remains indifferent to the outcome of random variable $X(w)$' (see Ben-Tal and Teboulle [7]), (2) the Shortfall risk - the minimum amount of additional resources needed to make the expected utility of a decision maker from his portfolio nonnegative, and (3) the Optimized Certainty Equivalent - representing the optimal allocation of $X(w)$ between present and future consumption.

Yet another type of risk measures are the tail-oriented risk measures. The most popular of them, the Value-at-Risk, represents the negative of the left α-th quantile of the distribution of $X(w)$. A constraint imposed on the VaR is equivalent to a chance constraint, as noted in Nemirovski and Shapiro [39] and used in Bertsimas et al. [13], Jiang and Guan [32], Hu and Hong [29], and Hu et al. [30]. However, efficient optimization of the VaR is difficult unless distributional assumptions are made. This problem can be mitigated by using instead of it a special case of the Optimized Certainty Equivalent, the Conditional Value-at-Risk, which represents the negative of the average of the worst $100\alpha\%$ outcomes of $X(w)$. Nemirovski and Shapiro [39] have shown the CVaR to be the best conservative convex approximation of the VaR. In Table 2.2 we use the formulation of CVaR adopted from Rockafellar and Uryasev [43]. Another advantage of the CVaR compared to the VaR is that it provides information about the mean of the least $100\alpha\%$ worst positions, instead of only the largest of them as is the case for the VaR.

Recently, an upper bound on both the VaR and the CVaR has been proposed by Ahmadi-Javid [2] - the Entropic Value-at-Risk. Its definition in Table 2.2 requires a separate comment since it does not involve p. Instead, EVaR is defined as a supremum over probability vectors $\tilde{p}$ in $\mathcal{P}_q$, constructed around a vector q. In this case the vector q shall be subject to uncertainty within a set $\mathcal{Q}$ - see the 'combined uncertainty set' in Table 2.3. We have chosen this formulation to make the notation of the corresponding function $f(p, w)$ (given in Section 2.5) consistent with the terminology of Theorem 2.1.

Table 2.2 includes also general risk measure classes whose definitions are based on axioms that risk measures should satisfy. In particular, these are the spectral, coherent, and convex risk measures, which are increasing classes in the sense of set inclusion. For a discussion of the differences between these three types of risk measures we refer the reader to Acerbi [1] and Föllmer and Schied [24].

Some of the measures in Table 2.2 are specific cases of the other ones: for instance, the CVaR is an example of an Optimized Certainty Equivalent. Nevertheless, a distinction has been made because of the popularity of the use of the specific cases. Also, some results can be obtained only for specific cases and it is important to state why this is so, and what the consequences are for practical applications.

For a further reference, each of the papers mentioned in Table 2.1 includes also a discussion of the applications of the relevant risk measures. For a broader overview of risk measurement and possible choices for risk measures we refer the reader to Embrechts et al. [21], Dowd [19], and Rockafellar and Royset [42].

2.4.2 Uncertainty sets for the probabilities

Distributional uncertainty has been studied up to now mostly in finance (see, e.g. Calafiore [16] and other references in Table 2.1), insurance (see, e.g., Klugman et al. [34]), economics (denoted in this context as ambiguity, see, e.g., Epstein [22]), and machine learning (see, e.g., Gotoh and Uryasev [27]).

Since some of the uncertainty sets are constructed using information on the outcomes of an underlying random vector, we assume that $X_n(w)$ corresponds to the outcome $Y^n \in \mathbb{R}^{M_Y}$ of some underlying random vector Y. Table 2.3 presents the uncertainty sets for the discrete probabilities studied in this paper.

We follow the view, motivated in Rühlicke [45], that the formulation of an uncertainty set for a probability distribution should be supported by results in statistics. For that reason, most of the sets in Table 2.3, including the Pearson, likelihood ratio, Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, or Kuiper sets, are constructed using goodness-of-fit test statistics with the corresponding names. Goodness-of-fit tests are typically used to test the hypothesis that a random sample with empirical probability distribution p has been sampled from an underlying distribution q. For that purpose, a test statistic is computed and compared to the critical value, based on the test type, sample size, and the chosen confidence level (see, for example, Thas [48]).

Example 2.2 *The test statistic corresponding to the Pearson test is given by*

$$D = \sum_{n \in \mathcal{N}} \frac{(p_n - q_n)^2}{q_n},$$

where p_n and q_n, are the empirical and the postulated (tested) probability of the n-th outcome, $n \in \mathcal{N}$, respectively. If, at the assumed confidence level, it holds that $D > D_0$, where D_0 is the critical level at the given confidence level, the hypothesis that the given random sample comes from the underlying distribution q, is rejected. Otherwise, there is no evidence to reject this hypothesis.

Table 2.2 – Risk measures analyzed in this paper. The term $\mathbb{E}^p$ denotes expectation with respect to the probability measure induced by the vector p and $G_{X(w)}$ denotes the distribution function of the random variable $X(w)$ as induced by p. We define the left α-quantile of a distribution of $X(w)$ as $G^{-1}_{X(w)}(\alpha) = \inf\{\kappa \in \mathbb{R} : \mathbb{P}(X(w) \leq \kappa) \geq \alpha\}$. The utility functions $u(\cdot)$ are assumed to be defined on the entire real line.

Risk measure	**Formulation** $F(p, w)$
Negative mean return	$-\mathbb{E}^p(X(w))$
Standard deviation less the mean	$\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2} - \alpha\mathbb{E}^p(X(w)), \quad \alpha \geq 0$
Standard deviation	$\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2}$
Variance less the mean	$\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2 - \alpha\mathbb{E}^p(X(w)), \quad \alpha \geq 0$
Variance	$\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2$
Mean absolute deviation from the median	$\mathbb{E}^p\left\lvert X(w) - G^{-1}_{X(w)}(0.5)\right\rvert$
Sharpe ratio	$\frac{-\mathbb{E}^p(X(w))}{\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2}}$
Lower partial moment	$\mathbb{E}^p(\max\{0, \bar{\kappa} - X(w)\}^\alpha)$ $\alpha = 1, 2, \quad \bar{\kappa}$ - any value
Certainty Equivalent (CE)	$-u^{-1}(\mathbb{E}^p u(X(w)))$ $u(\cdot)$ concave, invertible, with $-\frac{u'(t)}{u''(t)}$ concave
Shortfall risk	$\inf\{\kappa \in \mathbb{R} : \mathbb{E}^p(u(X(w) + \kappa) \geq 0\}$ $u(\cdot)$ concave
Optimized Certainty Equivalent (OCE)	$\inf_{\kappa \in \mathbb{R}} -\kappa - \mathbb{E}^p(u(X(w) - \kappa))$, $u(\cdot)$ concave, nondecreasing
Value-at-Risk (VaR)	$-G^{-1}_{X(w)}(\alpha), \quad 0 < \alpha < 1$
Conditional Value-at-Risk (CVaR)	$\inf_{\kappa \in \mathbb{R}} -\kappa - \mathbb{E}^p\left(\frac{1}{\alpha}\min\{X(w) - \kappa, 0\}\right), \quad 0 < \alpha < 1$
Entropic Value-at-Risk (EVaR) (see comments in Section 2.4.1)	$\sup_{\tilde{p} \in \mathcal{P}_q} \mathbb{E}^{\tilde{p}}(-X(w)), \quad 0 < \alpha < 1$ $\mathcal{P}_q = \left\{\tilde{p} : \tilde{p} \geq 0, 1^T\tilde{p} = 1, \sum_{n \in \mathcal{N}} \tilde{p}_n \log\left(\frac{\tilde{p}_n}{q_n}\right) \leq -\log\alpha\right\}$
Mean deviation from the mean	$\mathbb{E}^p\lvert X(w) - \mathbb{E}^p(X(w))\rvert$
Distortion risk measures	$\int_0^{+\infty} g\left(1 - G_{X(w)}(t)\right)dt$, $X(w)$ nonnegative, $g : [0,1] \to [0,1]$
Coherent risk measures	$\sup_{\tilde{p} \in \mathcal{C}} \mathbb{E}^{\tilde{p}}(-X(w)), \quad \mathcal{C}$ - set of probability vectors
Spectral risk measures	$-\int_0^1 G^{-1}_{X(w)}(t)\psi(t)dt$, $\psi(\cdot)$ nonnegative, non-increasing, right-continuous, integrable

For examples of φ-divergences we refer the reader to Table 2.5 in Appendix 2.C.1. The Pearson and likelihood ratio sets are specific cases of the φ-divergence set (obtained by choosing the Kullback-Leibler or the modified χ^2 divergences, respectively), but have been distinguished here because of their popularity. The likelihood ratio set, having some computational and statistical advantages due to its relation to information theory, is a common choice in studies considering distributional uncertainty, see, for example, Hu and Hong [29], Hu et al. [30], and Jiang and Guan [32].

The Wasserstein set definition, using the Wasserstein (Kantorovich) distance between distribution vectors p and q, deserves a separate explanation. The distance between p and q, defined with the use of the inf term in Table 2.3, can be interpreted as a minimum transport cost of the probability mass from vector p (supply) to vector q (demand), where the unit cost between the i-th cell of p and the j-th cell of q is equal to $\|Y^i - Y^j\|^d$, where Y^i is the i-th observation of the underlying random variable Y, as specified in Section 2.3. This type of uncertainty is studied extensively in a robust setting in Wozabal [51] and the statistical advantages of its use are motivated in Rühlicke [45].

A separate explanation is also needed for the 'combined uncertainty set'. This class of uncertainty sets has been introduced here to derive the tractable robust counterpart of a constraint on the Entropic Value-at-Risk. Its definition in Table 2.3 says that $\mathcal{P}^{\mathrm{C}}$ has a two-stage structure. First, the vector p belongs to a set $\mathcal{P}_q$ centered around a vector q. Then, the vector q is uncertain itself and belongs to a set $\mathcal{Q}$ defined using Q convex inequalities. In this paper we shall assume that $\mathcal{P}_q$ is defined as a φ-divergence set around q, as in the first row of Table 2.3.

Some of the formulations in Table 2.3 include explicitly both the vectors p and q and the others only the vector p (with an abbreviation 'emp' next to set name in the last column of Table 2.3). The former case corresponds to the situation where the uncertainty set for p is defined with reference to a nominal distribution q that in principle can be chosen arbitrarily. In such a case, a typical choice for q will be the empirical distribution. The latter case corresponds to goodness-of-fit tests for one-dimensional random samples $Y^1 \leq Y^2 \leq \ldots \leq Y^N$. There, the nominal measure q is implicitly defined by the empirical distribution of the sample at hand and cannot be chosen arbitrarily.

This does not mean that one can use such uncertainty sets only for the case where Y is one-dimensional. Under the assumption that the dependence structure of the marginal distributions is unknown, a simple goodness-of-fit test (and hence, an uncertainty set) can be constructed by applying a given goodness-of-fit test for each marginal distribution, with modified confidence level. Such an approach is applied, for example, by Bertsimas et al. [14], including some of the uncertainty sets considered in Table 2.3.

For other dependence structures between the marginal distributions, however, construction of credible goodness-of-fit tests is still an area of research. Bertsimas et al. [13] show how to use the Kolmogorov-Smirnov test for each of the marginal distributions, under the assumption of their independence, to obtain approximations of constraints on the VaR.

If needed, continuous probability distributions can be transformed to discrete ones to fit into the framework of our survey, e.g., by dividing the support of a random variable into cells and using the per-cell probabilities. However, one has to be aware of potential shortcomings of such an approach in statistical testing, see, for example, Thas [48].

The choice of an uncertainty set for a particular application is related to the properties of the problem at hand, such as the dimensionality of data, number of observations, and properties of the goodness-of-fit test/probability metric on which a given set is based. Similar to the case of risk measures, the choice of a specific uncertainty set type has its implications for the complexity of the reformulated problem, see Table 2.1. For an overview and further discussion of statistical tests and probability metrics we refer the reader to Gibbs and Su [26] and Thas [48].

We now verify under which conditions the uncertainty sets listed in Table 2.3 satisfy condition (2.8) so that we can apply Theorem 2.1 to derive the tractable robust counterparts. For the set definitions with no q involved, condition (2.8) holds. For sets involving a vector q, (2.8) holds if we assume that (i) $q \in \mathcal{P}$, and (ii) $q \in \mathbb{R}^N_{++}$ (since then the sets $\mathrm{ri}(\mathcal{P})$ and $\mathbb{R}^N_{++}$ obviously have a point q in common). Condition (i) is reasonable since the best estimate of the uncertain probability vector should also belong to the uncertainty set. For (ii), if there were a scenario j such that $q_j = 0$, then such a scenario should not be taken into account, since an estimated probability equal to zero means that the scenario is empirically irrelevant. Thus, the assumptions (i) and (ii) can be expected to hold in applications of our methodology.

2.5 Conjugates of the risk measures

In Table 2.2 each risk measure corresponds to a specific function $F(p, w)$, defining its value for w under the probability distribution induced by the vector p. However, not for all cases it is possible to apply Theorem 2.1 using in the forms presented in Table 2.2. For such cases, we may need an equivalent formulation of the risk measure using some new function $f(p, w)$, for which the assumptions of Theorem 2.1 are satisfied.

In this section we give the results on concave conjugates $f_*(v, w)$ of such relevant functions $f(p, w)$ corresponding to the risk measures from Table 2.2. For some cases we take $f(p, w) = F(p, w)$. For others, such as the Optimized Certainty Equivalent or the variance, $F(p, w)$ is reformulated to an equivalent form using a new function

Table 2.3 – Uncertainty set formulations for the probabilities vector p. In each case we assume that $p \geq 0, 1^T p = 1$ hold.

Set type	**Formulation**	**Symbol**
φ-divergence	$\sum_{n\in\mathcal{N}} q_n \varphi\left(\frac{p_n}{q_n}\right) \leq \rho$	$\mathcal{P}_q^{\varphi}$
Pearson (χ^2)	$\sum_{n\in\mathcal{N}} \frac{(p_n-q_n)^2}{q_n} \leq \rho$	$\mathcal{P}_q^{\mathrm{P}}$
Likelihood ratio (G)	$\sum_{n\in\mathcal{N}} q_n \log\left(\frac{p_n}{q_n}\right) \leq \rho$	$\mathcal{P}_q^{\mathrm{LR}}$
Kolmogorov-Smirnov	$\max_{n\in\mathcal{N}} \left\lvert p^T 1^n - q^T 1^n \right\rvert \leq \rho$	$\mathcal{P}_q^{\mathrm{KS}}$
Wasserstein (Kantorovich)	$\inf_{\substack{K:K_{ij}\geq 0,\forall i,j \\ K1=q, K^T 1=p}} \left(\sum_{i,j\in\mathcal{N}} K_{ij} \lVert Y^i - Y^j \rVert^d \right) \leq \rho, \quad d \geq 1$	$\mathcal{P}_q^{\mathrm{W}}$
Combined set	$p \in \mathcal{P}_q, \quad q \in \mathcal{Q} = \{q : \quad h_i(q) \leq 0, \quad i = 1, ..., Q\}$	$\mathcal{P}^{\mathrm{C}}$
Anderson-Darling	$-N - \sum_{n\in\mathcal{N}} \frac{2n-1}{N} \left(\log\left(p^T 1^n\right) + \log\left(p^T 1^{-n}\right)\right) \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{AD}}$
Cramer-von Mises	$\frac{1}{12N} + \sum_{n\in\mathcal{N}} \left(\frac{2n-1}{2N} - p^T 1^n\right)^2 \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{CvM}}$
Watson	$\frac{1}{12N} + \sum_{n\in\mathcal{N}} \left(\frac{2n-1}{2N} - p^T 1^n\right)^2 - N\left(\frac{1}{N}\sum_{n\in\mathcal{N}} p^T 1^n - \frac{1}{2}\right)^2 \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{Wa}}$
Kuiper	$\max_{n\in\mathcal{N}} \left(\frac{n}{N} - p^T 1^n\right) + \max_{n\in\mathcal{N}} \left(p^T 1^{n-1} - \frac{n-1}{N}\right) \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{K}}$

$f(p, w)$ linear in p:

$$f(p, w) = Z_0 + \sum_{n\in\mathcal{N}} p_n Z_n(w),$$

for appropriate Z_0 and $Z_n(w)$. Linearity in p is a desirable property since then the conjugate $f_*(v, w)$ follows directly from (2.4):

$$f_*(v, w) = \begin{cases} -Z_0 & \text{if } Z_n(w) \leq v_n, \quad \forall n \in \mathcal{N} \\ -\infty & \text{otherwise.} \end{cases} \tag{2.10}$$

Derivations for the risk measures where even the $f(p, w)$ is nonlinear in p are given in Appendix 2.B. The remainder of this section distinguishes three cases, depending on the type of the functions $F(p, w)$ and $f(p, w)$: (1) when $F(p, w) \equiv f(p, w)$ is linear in p, (2) when $F(p, w)$ is nonlinear in p, but $f(p, w)$ is linear in p, and (3) when both $F(p, w)$ and $f(p, w)$ are nonlinear in p. For each conjugate function we give the complexity of the system of inequalities in the formulation under the condition that each $X_n(\cdot)$ is linear.

2.5.1 $F(p, w)$ linear in p

In this subsection we analyze the risk measures for which $F(p, w) \equiv f(p, w)$ is linear in p.

Negative mean return. For the negative mean return the function is:

$$f(p,w) = F(p,w) = \sum_{n\in\mathcal{N}} p_n \left(-X_n(w)\right).$$

Its concave conjugate is given by formula (2.10) with $Z_0 = 0$ and $Z_n(w) = -X_n(w)$. If each $X_n(\cdot)$ is linear, the inequalities in this formulation are linear in w.

Shortfall risk. In case of the Shortfall risk the constraint itself is imposed on the variable κ. The constraint to be reformulated is $\mathbb{E}^p u(X(w)+\kappa) \geq 0$ or, equivalently:

$$-\mathbb{E}^p u(X(w)+\kappa) \leq 0, \quad \forall p \in \mathcal{P}.$$

The function $f(p,w)$ we take is:

$$f(p,w) = -\sum_{n\in\mathcal{N}} p_n u(X_n(w)+\kappa).$$

Its conjugate is given by (2.10) with $Z_0 = 0$ and $Z_n(w) = -u(X_n(w)+\kappa)$. If each $X_n(\cdot)$ is linear, then, due to the concavity of $u(\cdot)$, the inequalities included in this formulation are convex in the decision variables.

Lower partial moment. In this case the function is:

$$f(p,w) = F(p,w) = \sum_{n\in\mathcal{N}} p_n \max\left\{0, \bar{\kappa} - X_n(w)\right\}^{\alpha}.$$

Its conjugate is given by (2.10) with $Z_0 = 0$ and $Z_n(w) = \max\left\{0, \bar{\kappa} - X_n(w)\right\}^{\alpha}$. If each $X_n(\cdot)$ is linear, then for $\alpha = 1$ the inequalities involved are linear, and for $\alpha = 2$ they are convex quadratic in the decision variables.

2.5.2 $F(p,w)$ nonlinear in p and $f(p,w)$ linear in p

In this subsection we analyze the risk measures for which $F(p,w)$ is nonlinear in p but $f(p,w)$ is linear in p.

Optimized Certainty Equivalent. For a constraint on the OCE, the constraint is:

$$F(p,w) = \inf_{\kappa\in\mathbb{R}} \left\{-\kappa - \sum_{n\in\mathcal{N}} p_n(u(X_n(w)-\kappa))\right\} \leq \beta, \quad \forall p \in \mathcal{P}. \tag{2.11}$$

Due to Lemma 2.2 (see Appendix 2.B.1), for continuous and finite-valued functions $u(\cdot)$ and compact sets $\mathcal{P}$ (being the uncertainty set for probabilities in our case) it holds that

$$\sup_{p\in\mathcal{P}} \inf_{\kappa\in\mathbb{R}} \left\{-\kappa - \sum_{n\in\mathcal{N}} p_n(u(X_n(w)-\kappa))\right\} = \inf_{\kappa\in\mathbb{R}} \sup_{p\in\mathcal{P}} \left\{-\kappa - \sum_{n\in\mathcal{N}} p_n(u(X_n(w)-\kappa))\right\}.$$

Using this result, the inf term in (2.11) can be removed, and the following constraint, with κ as a variable, is equivalent to (2.11):

$$f(p,w) = -\kappa - \sum_{n\in\mathcal{N}} p_n(u(X_n(w)-\kappa)) \leq \beta, \quad \forall p \in \mathcal{P}.$$

This formulation is already in the form of Theorem 2.1 and the concave conjugate of $f(p,w)$ with respect to its first argument is given by (2.10) with $Z_0 = -\kappa$ and $Z_n(w) = -u(X_n(w)-\kappa)$. If each $X_n(\cdot)$ is linear, then this formulation involves convex inequalities in the decision variables. For the Conditional Value-at-Risk, as a special case of the OCE, we have $Z_0 = -\kappa$ and $Z_n(w) = -\frac{1}{\alpha}\min\{X_n(w)-\kappa, 0\}$. If each $X_n(\cdot)$ is linear, the inequalities included in this formulation are representable as a system of linear inequalities in the decision variables.

Certainty Equivalent. For general $u(\cdot)$ the formulation of a conjugate function would involve inequalities that are nonconvex in the decision variables. If one assumes that β is a fixed number, then a more tractable way to include a constraint on the CE:

$$F(p,w) = -u^{-1}\left(\sum_{n\in\mathcal{N}} p_n u(X(w))\right) \leq \beta, \quad \forall p \in \mathcal{P}$$

is to multiply both sides by -1, then apply the function $u(\cdot)$ to both sides to arrive at an equivalent constraint

$$\tilde{F}(p,w) = -\sum_{n\in\mathcal{N}} p_n u(X(w)) \leq -u(-\beta), \quad \forall p \in \mathcal{P}.$$

This constraint is of the same type as the robust constraint for the Shortfall risk. Therefore, the result for Shortfall risk can be used to obtain the relevant concave conjugate. In this case one cannot combine the CE with other risk measures via using the β as a variable.

Mean absolute deviation from the median. The constraint for this risk measure is given by:

$$F(p,w) = \sum_{n\in\mathcal{N}} p_n \left|X_n(w) - G^{-1}_{X(w)}(0.5)\right| \leq \beta, \quad \forall p \in \mathcal{P}.$$

Because of the median, $G^{-1}_{X(w)}(0.5)$, the function above is nonlinear in p and its concavity status is difficult to determine. However, we have:

$$F(p,w) = \sum_{n\in\mathcal{N}} p_n \left|X_n(w) - G^{-1}_{X(w)}(0.5)\right| = \inf_{\kappa\in\mathbb{R}} \sum_{n\in\mathcal{N}} p_n \left|X_n(w)-\kappa\right|.$$

This result is obtained by considering the impact of changing κ on the value of the sum on the right hand-side, separately for the cases $\kappa > G^{-1}_{X(w)}(0.5)$ and $\kappa < G^{-1}_{X(w)}(0.5)$.

By formulating $F(p,w)$ as an infimum over linear functions in p, we immediately know that it is also concave in p. The conditions of Lemma 2.2 (see Appendix 2.B.1)

are therefore satisfied so that, similar to the Optimized Certainty Equivalent, we can remove the inf term to study equivalently the robust constraint on the following function:

$$f(p, w) = \sum_{n \in \mathcal{N}} p_n \left| X_n(w) - \kappa \right|,$$

where κ is a variable. Its conjugate is given by (2.10) with $Z_0 = 0$ and $Z_n(w) = |X_n(w) - \kappa|$. If each $X_n(\cdot)$ is linear, the inequalities included in the formulation above are representable as a system of linear inequalities in the decision variables.

Variance less the mean. The constraint for this risk measure is given by:

$$F(p, w) = \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w) \leq \beta, \quad \forall p \in \mathcal{P}.$$

Even though this formulation is concave in p, the results obtained in [10] for the variance in this form are difficult to implement. We propose to use, similar to the case of mean absolute deviation from the median, the following, well-known fact:

$$\begin{aligned} F(p, w) &= \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w) \\ &= \inf_{\kappa \in \mathbb{R}} \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \kappa \right)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w). \end{aligned} \tag{2.12}$$

Indeed, the minimized expression is strictly convex in κ. Deriving the first-order optimality condition results in $\kappa = \sum_{n \in \mathcal{N}} p_n X_n(w)$.

The conditions of Lemma 2.2 (see Appendix 2.B.1) are satisfied, thus we can remove the inf term to study equivalently the robust constraint on the following function:

$$f(p, w) = \sum_{n \in \mathcal{N}} p_n \left(\left(X_n(w) - \kappa \right)^2 - \alpha X_n(w) \right).$$

Its concave conjugate is given by (2.10) with $Z_0 = 0$ and $Z_n(w) = \left(X_n(w) - \kappa \right)^2 - \alpha X_n(w)$. The result for the variance is obtained by setting $\alpha = 0$. If each $X_n(\cdot)$ is linear, then this formulation involves convex quadratic inequalities in the decision variables.

Entropic Value-at-Risk. A robust constraint on the EVaR is given by

$$F(q, w) = \sup_{\tilde{p} \in \mathcal{P}_q} \mathbb{E}^{\tilde{p}}(-X(w)) \leq \beta, \quad \forall q \in \mathcal{Q}$$

with

$$\mathcal{P}_q = \left\{ \tilde{p} : \tilde{p} \geq 0, \quad 1^T \tilde{p} = 1, \quad \sum_{n \in \mathcal{N}} \tilde{p}_n \log \left(\frac{\tilde{p}_n}{q_n} \right) \leq -\log \alpha \right\},$$

and $\mathcal{Q}$ defined as in Table 2.3. The derivation of the concave conjugate with such a definition is troublesome since the function $F(q, w)$ is formulated as a supremum.

Because of this we introduce the notion of a combined uncertainty set to include the formulations of $\mathcal{P}_q$ and $\mathcal{Q}$ in the definition of a joint uncertainty set for (p, q) and to construct a relevant matrix A^{C}. Then, the robust constraint on the EVaR is:

$$\begin{aligned} f(p,w) &= \sum_{n\in\mathcal{N}} p_n\left(-X(w)\right) \leq \beta, \quad p = A^{\mathrm{C}}p', \quad A^{\mathrm{C}} = [I|0_{N\times N}], \\ &\forall p' \in \left\{ p' = \begin{bmatrix} \overline{p} \\ q \end{bmatrix} : p' \geq 0, \quad 1^T\overline{p} = 1, \quad \sum_{n\in\mathcal{N}} \overline{p}_n \log\left(\frac{\overline{p}_n}{q_n}\right) \leq \rho, q \in \mathcal{Q} \right\}. \end{aligned} \tag{2.13}$$

The function $f(p, w)$ for which the concave conjugate is to be derived, is the same as for the negative mean return, for which (2.10) holds with $Z_n(w) = -X_n(w)$ and $Z_0 = 0$. What is left, is the derivation of the support function for the uncertainty set for p', which is done in Section 2.6. The approach developed here for the EVaR can also be applied to other types of uncertainty sets $\mathcal{P}_q$.

2.5.3 Both $F(p, w)$ and $f(p, w)$ nonlinear in p

In this subsection we analyze the risk measures for which both $F(p, w)$ and $f(p, w)$ are nonlinear in p.

Standard deviation less the mean. The constraint on this risk measure is given by:

$$F(p,w) = \sqrt{\sum_{n\in\mathcal{N}} p_n \left(X_n(w) - \sum_{n'\in\mathcal{N}} p_{n'}X_{n'}(w)\right)^2} - \alpha \sum_{n\in\mathcal{N}} p_n X_n(w) \leq \beta, \quad \forall p \in \mathcal{P}.$$

The function $F(p, w)$ is nonlinear in p and a derivation of its conjugate would be troublesome. We use the fact that:

$$F(p,w) = \inf_{\kappa\in\mathbb{R}} \sqrt{\sum_{n\in\mathcal{N}} p_n(X_n(w) - \kappa)^2} - \alpha \sum_{n\in\mathcal{N}} p_n X_n(w).$$

This formulation follows for the same reason as in the case of variance, since the minimized expression is an increasing function (square root) of the variance. The conditions of Lemma 2.2 (see Appendix 2.B.1) are satisfied and, similar to the Optimized Certainty Equivalent, one can remove the inf term to reformulate equivalently the robust constraint on the following function:

$$f(p,w) = \sqrt{\sum_{n\in\mathcal{N}} p_n(X_n(w) - \kappa)^2} - \alpha \sum_{n\in\mathcal{N}} p_n X_n(w).$$

The function $f(p, w)$ is concave in p and we can use Theorem 2.1. The conjugate of $f(p, w)$ is equal to:

$$\begin{aligned} f_*(v, w) = \sup_{y} \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \left\| \begin{bmatrix} X_n(w) - \kappa \\ \left(\frac{v_n + \alpha X_n(w) - y}{2}\right) \end{bmatrix} \right\|_2 \leq \tfrac{v_n + \alpha X_n(w) + y}{2}, \quad \forall n \in \mathcal{N} \\ & v_n + \alpha X_n(w) \geq 0, \quad \forall n \in \mathcal{N} \\ & y \geq 0. \end{aligned} \tag{2.14}$$

The derivation can be found in Appendix 2.B. If each $X_n(\cdot)$ is linear, the above formulation involves second-order conic inequalities in the decision variables. The result for the standard deviation is obtained by setting $\alpha = 0$. One can note that the sup-formulation of the conjugate function fits well into (2.9) since the conjugate function appears there with a negative sign. Thus, one can omit the sup symbol after inserting (2.14) into (2.9) and still have an equivalent constraint.

Sharpe ratio. A robust constraint on the Sharpe ratio risk measure is:

$$F(p, w) = \frac{-\sum\limits_{n \in \mathcal{N}} p_n \left(X_n(w)\right)}{\sqrt{\sum\limits_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum\limits_{n' \in \mathcal{N}} p_{n'} X_{n'}(w)\right)^2}} \leq \beta, \quad \forall p \in \mathcal{P}.$$

The left-hand side function is neither convex, nor concave in the probabilities and we did not find a more tractable function $f(p, w)$ for it. If one assumes that β is a fixed number, then the constraint can be reformulated equivalently to:

$$\sqrt{\sum_{n \in \mathcal{N}} p_n (X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w))^2} - \frac{1}{\beta} \sum_{n \in \mathcal{N}} p_n \left(X_n(w)\right) \leq 0, \quad \forall p \in \mathcal{P}.$$

This constraint is equivalent to a robust constraint on the standard deviation less the mean with $\alpha = 1/\beta$ and the right hand side equal to 0. Thus, the corresponding result can be used for the conjugate function. In this case one cannot combine the Sharpe ratio with other risk measures using β as a variable.

In the case of VaR we did not find a formulation of the risk measure that would allow us to find a closed-form concave conjugate. A similar situation occurred for the general distortion, spectral, and coherent risk measures. We found the structure of their definitions intractable unless, for example, a coherent risk measure can be analyzed using a combined uncertainty set, as in the case of EVaR. The mean absolute deviation from the mean is nonconvex and nonconcave in the probabilities and for that reason we could not obtain a closed-form or inf-form for its concave conjugate.

2.6 Support functions of the uncertainty sets

In this section, the formulations of the support functions are given for the sets $\mathcal{U}$ corresponding to the uncertainty sets listed in Table 2.3. Our results of this section utilize heavily the property of (2.9), where it is sufficient to have the support function formulated as an infimum. Then, the inf of the support function symbol can be dropped after inserting the expression for the support function into (2.9), as explained in Remark 2.2. We need this property as most of the support functions of the uncertainty sets have been obtained using the following lemma, taken from [10]:

Lemma 2.1 *Let $Z \subset \mathbb{R}^L$ be of the form $Z = \{\zeta : \quad h_i(\zeta) \leq 0, \quad i = 1, ..., H\}$, where the $h_i(\cdot)$ is convex for each i. If it holds that $\cap_{i=1}^{H} ri(domh_i) \neq \emptyset$, then:*

$$\delta^*(v|Z) = \inf_{u \geq 0} \left\{ \sum_{i=1}^{H} u_i h_i^* \left(\frac{v^i}{u_i} \right) \middle| \sum_{i=1}^{H} v^i = v \right\},$$

where we define $0h_i^*(v^i/0) = \lim_{u_i \to 0^+} u_i h_i^*(v^i/u_i)$.

For each of the support functions we proceed in the same way. First, we give the necessary parameters, assuming that $A = I$ and $\mathcal{P} = \mathcal{U}$, unless stated otherwise. Then the support function follows, referring to Appendix 2.C for the derivations.

φ-divergence functions. For the uncertainty set defined using the φ-divergence the support function is (see Appendix 2.C.2 for a derivation):

$$\delta^*\left(v \middle| \mathcal{P}_q^{\varphi}\right) = \inf_{u \geq 0, \eta} \left\{ \eta + u\rho + u \sum_{n \in \mathcal{N}} q_n \varphi^* \left(\frac{v_n - \eta}{u} \right) \right\}. \tag{2.15}$$

This result has also been obtained in [9]. In the general case the right-hand side expression between the brackets is a nonlinear convex function of the decision variables. However, for specific choices (see Table 2.5 in Appendix 2.C.1) it can have more tractable forms - for instance, for the Variation distance it is linear. Result (2.15) holds also for the Pearson and likelihood ratio sets since they are specific cases of the φ-divergence set.

Kolmogorov-Smirnov. For an uncertainty set defined using the Kolmogorov-Smirnov test we take a matrix $D \in \mathbb{R}^{(2N+2)\times N}$ and a vector $d \in \mathbb{R}^{2N+2}$ whose components are:

$$\begin{array}{llll}
D_{1n} = 1, & d_1 = 1, & \forall n \in \mathcal{N} & \\
D_{2n} = -1, & d_2 = -1, & \forall n \in \mathcal{N} & \\
D_{2+n,i} = 1, & d_{2+n} = \rho + q^T 1^n, & \forall i \leq n, & n \in \mathcal{N} \\
D_{2+N+n,i} = -1, & d_{2+N+n} = \rho - q^T 1^n, & \forall i \leq n, & n \in \mathcal{N},
\end{array}$$

with the other components equal to 0. Under such a parametrization, the support function is equal to (see Appendix 2.C.3 for a derivation):

$$\begin{aligned} \delta^*\left(v \,\middle|\, \mathcal{P}^{\mathrm{KS}}\right) = \inf_u \quad & u^T d \\ \text{s.t.} \quad & v \leq D^T u \\ & u \geq 0. \end{aligned} \tag{2.16}$$

The optimization problem in (2.16) is linear.

Wasserstein. For an uncertainty set defined using the Wasserstein distance we take $A^{\mathrm{W}} = [I\,|0_{N\times N^2}]$. This choice is motivated in the derivation in Appendix 2.C.4. Also, a matrix $D \in \mathbb{R}^{(4N+3)\times(N^2+N)}$ and a vector $d \in \mathbb{R}^{4N+3}$ are needed, whose components are:

$$\begin{aligned}
&D_{1n} = 1, && d_1 = 1, && \forall n \in \mathcal{N} \\
&D_{2n} = -1, && d_2 = -1, && \forall n \in \mathcal{N} \\
&D_{3,Ni+n} = \|Y_i - Y_n\|^d, && d_3 = \rho, && \forall i, n \in \mathcal{N} \\
&D_{3+n,n} = -1, && D_{3+n,Nn+i} = 1, && \forall i, n \in \mathcal{N} \\
&D_{3+N+n,n} = 1, && D_{3+N+n,Nn+i} = -1, && \forall i, n \in \mathcal{N} \\
&D_{3+2N+n,Ni+n} = 1, && d_{3+2N+n} = q_n, && \forall i, n \in \mathcal{N} \\
&D_{3+3N+n,Ni+n} = -1, && d_{3+3N+n} = -q_n, && \forall i, n \in \mathcal{N},
\end{aligned}$$

with all other components of D and d equal to 0. The corresponding support function is equal to (see Appendix 2.C.4 for a derivation):

$$\begin{aligned} \delta^*\left(\left(A^{\mathrm{W}}\right)^T v \,\middle|\, \mathcal{U}_q^{\mathrm{W}}\right) = \inf_u \quad & u^T d \\ \text{s.t.} \quad & (A^{\mathrm{W}})^T v \leq D^T u \\ & u \geq 0. \end{aligned} \tag{2.17}$$

The optimization problem in (2.17) is linear.

Combined set. We assume that the uncertainty set $\mathcal{P}_q$ is defined as a φ-divergence set around q (being the Kullback-Leibler divergence for the EVaR). We take a matrix $A^{\mathrm{C}} = [I|0_{N\times N}]$, motivated in the corresponding section of Appendix 2.C.5. The

support function is equal to (see Appendix 2.C.5 for a derivation):

$$\begin{aligned}
\delta^*\left(\left(A^{\mathrm{C}}\right)^T v \middle| \mathcal{U}^{\mathrm{C}}\right) = \inf_{\substack{\{u_i, v^i\},\\ i=1,\ldots,Q+3}} \quad & u_1 - u_2 + u_3\rho + \sum_{i=1}^{Q} u_{i+3} h_i^*\left(\tfrac{v_{N+1:2N}^{i+3}}{u_{i+3}}\right) \\
\text{s.t.} \quad & v_1 \leq u_1 1 \\
& v_{1:N}^2 \leq -u_2 1 \\
& v_{N+1:2N}^i = 0, \quad i = 1,2,3 \\
& v_{1:N}^i = 0, \quad i = 4,\ldots,Q+3 \\
& v_{N+n}^3 + u_3 \varphi^*\left(\tfrac{v_n^3}{u_3}\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=1}^{Q+3} v^i = \left(A^{\mathrm{C}}\right)^T v \\
& u_i \geq 0, \quad i = 1,\ldots,Q+3.
\end{aligned} \tag{2.18}$$

For all φ-divergence functions listed in Table 2.5 the optimization problem in (2.18) is convex. If the φ-divergence is the Variation distance or the modified χ^2 distance and the functions $h_i(\cdot)$ are all linear or convex quadratic, then the optimization problem in (2.18) is linear or convex quadratic, respectively.

Anderson-Darling. For an uncertainty set defined using the Anderson-Darling test the support function $\delta^*\left(v \middle| \mathcal{P}_{\mathrm{emp}}^{\mathrm{AD}}\right)$ is equal to (see Appendix 2.C.6 for a derivation):

$$\begin{aligned}
\inf_{\eta, u, \{w^{n+}, w^{n-}\}, n \in \mathcal{N}} \quad & -\sum_{n \in \mathcal{N}} \tfrac{(2n-1)u}{N}\left[2 + \log\left(\tfrac{-Nz_n^+}{(2n-1)u}\right) + \log\left(\tfrac{-Nz_n^-}{(2n-1)u}\right)\right] \\
& + u\left(\rho + N\right) + \eta \\
\text{s.t.} \quad & v \leq \sum_{n \in \mathcal{N}} \left(z_n^+ 1^n + z_n^- 1^{-n}\right) + \eta 1 \\
& z_n^+, z_n^- \leq 0 \quad \forall n \in \mathcal{N} \\
& u \geq 0.
\end{aligned} \tag{2.19}$$

This result has also been obtained in [10]. The optimization problem in (2.19) is convex.

Cramer-von Mises. For an uncertainty set defined using the Cramer-von Mises test we use the following parameters:

$$c = -\rho + \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N}\right)^2, \quad b = \begin{bmatrix} -2\sum_{j=1}^{N} \frac{2j-1}{N} \\ -2\sum_{j=2}^{N} \frac{2j-1}{N} \\ \vdots \\ -2\sum_{j=N}^{N} \frac{2j-1}{N} \end{bmatrix},$$

a matrix $E \in \mathbb{R}^{N\times N}$ such that $E_{ij} = N + 1 - \max\{i, j\}$ for $i, j \in \mathcal{N}$ and a unique matrix P such that $P^T P = E^{-1}$. With such a parametrization, the support function is equal to (see Appendix 2.C.7 for a derivation):

$$\begin{aligned}\delta^*\left(v \,\middle|\, \mathcal{P}^{\mathrm{CvM}}_{\mathrm{emp}}\right) = \inf_{z,t,\{u_i,v^i\},i=1,\ldots,3} \quad & u_1 - u_2 + \tfrac{1}{4}t - u_3 c \\ \text{s.t.} \quad & \left\| \begin{bmatrix} Pz \\ \frac{t-u_3}{2} \end{bmatrix} \right\|_2 \leq \tfrac{t+u_3}{2} \\ & z = u_3 b - v^3 \\ & u_1 - u_2 + v_n^3 - v_n \geq 0, \quad \forall n \in \mathcal{N} \\ & u_1, u_2, u_3 \geq 0.\end{aligned} \tag{2.20}$$

The optimization problem in (2.20) is convex quadratic.

Watson. For an uncertainty set defined using the Watson test we use the following parameters:

$$c = -\rho + \frac{1}{12N} + \sum_{n\in\mathcal{N}} \left(\frac{2n-1}{2N}\right)^2 - \frac{N}{4}, \quad b = \begin{bmatrix} -2\sum_{j=1}^{N} \frac{2j-1}{N} + N \\ -2\sum_{j=2}^{N} \frac{2j-1}{N} + (N-1) \\ \vdots \\ -2\sum_{j=N}^{N} \frac{2j-1}{N} + 1 \end{bmatrix},$$

a matrix $E \in \mathbb{R}^{N\times N}$ such that:

$$E_{i,j} = N + 1 - \max\{i, j\} - \frac{(N+1-i)(N+1-j)}{N}, \quad \forall i, j \in \mathcal{N},$$

and a matrix P such that $P^T P = E$. With such a parametrization, the support function is given by (see Appendix 2.C.8 for a derivation):

$$\begin{aligned}\delta^*\left(v \,\middle|\, \mathcal{P}^{\mathrm{Wa}}_{\mathrm{emp}}\right) = \inf_{\substack{z,t,\lambda,\{u_i,v^i\},\\ i=1,\ldots,3}} \quad & u_1 - u_2 + \tfrac{1}{4}t - u_3 c \\ \text{s.t.} \quad & \left\| \begin{bmatrix} Pz \\ \frac{t-u_3}{2} \end{bmatrix} \right\|_2 \leq \tfrac{t+u_3}{2} \\ & z = u_3 b - v^3 \\ & u_1 - u_2 + v_n^3 - v_n \geq 0, \quad n \in \mathcal{N} \\ & u_1, u_2, u_3, t \geq 0 \\ & E\lambda = z.\end{aligned} \tag{2.21}$$

The optimization problem in (2.21) is convex quadratic.

Kuiper. For the uncertainty set defined using the Kuiper test we take $A^{\mathrm{K}} = [I\,|0_{N\times 2}]$. Also, a matrix $D \in \mathbb{R}^{(2N+3)\times(N+2)}$ and a vector $d \in \mathbb{R}^{2N+3}$ are used, whose components are:

$$\begin{array}{llll}
D_{1,n} = 1, & d_1 = 1, & & \forall n \in \mathcal{N} \\
D_{2,n} = -1, & d_2 = -1, & & \forall n \in \mathcal{N} \\
D_{2+n,i} = -1, & D_{2+n,N+1} = -1, & d_{n+2} = -n/N, & \forall i \le n, n \in \mathcal{N} \\
D_{N+2+n,i} = 1, & D_{N+2+n,N+2} = -1, & d_{N+2+n} = (n-1)/N, & \forall i \le n-1, n \in \mathcal{N} \\
D_{2N+3,N+1} = 1, & D_{2N+3,N+2} = 1, & d_{2N+3} = \rho, &
\end{array}$$

with all other components of the matrix D and vector d equal to 0. Under such a parametrization, the support function is (see Appendix 2.C.9 for a derivation):

$$\begin{aligned}
\delta^*\left(\left(A^{\mathrm{K}}\right)^T v \middle| \mathcal{U}^{\mathrm{K}}_{\mathrm{emp}}\right) = \inf_u \quad & u^T d \\
\text{s.t.} \quad & \left(A^{\mathrm{K}}\right)^T v \le D^T u \\
& u \ge 0.
\end{aligned} \tag{2.22}$$

The optimization problem in (2.22) is linear.

2.7 Examples

In this section we present three examples of constraints or problems involving distributional uncertainty. The first example is a simple one where we demonstrate our unifying approach on a single constraint introduced earlier in the paper. The second example is in the field of finance and the third one is of industrial type - a data-driven antenna array design problem. The latter two examples are also studied numerically.

2.7.1 Standard deviation with φ-divergence uncertainty set

For a simple exposition of the advantages of our method, we shall derive a tractable counterpart of the constraint from Example 2.1 (see page 21), where the constraint is imposed on the standard deviation and the uncertainty set is defined by means of a φ-divergence uncertainty set:

$$\sqrt{\sum_{n\in\mathcal{N}} p_n \left(X_n(w) - \sum_{n'\in\mathcal{N}} p_{n'} X_{n'}(w)\right)^2} \le \beta, \quad \forall p \in \mathcal{P},$$

with

$$\mathcal{P} = \left\{p \ge 0 : \quad \sum_{n\in\mathcal{N}} p_n = 1, \quad \sum_{n\in\mathcal{N}} q_n \varphi\left(\frac{p_n}{q_n}\right) \le \rho\right\}.$$

In order to obtain a tractable robust counterpart of the form (2.9), we need to identify:

i the function $f(p, w)$ corresponding to the standard deviation and derive its conjugate,

ii the conjugate of the support function of the φ-divergence set.

The relevant component for (i) is (2.14), see page 36. For (ii), we refer to (2.15), page 37.

Inserting the two components into (2.9), we obtain an equivalent constraint:

$$\left(\begin{array}{ll} \inf\limits_{\eta,u} & \eta + u\rho + u\sum\limits_{n\in\mathcal{N}} q_n\varphi^*\left(\frac{v_n-\eta}{u}\right) \\ \text{s.t.} & u \geq 0 \\ & v_n + \alpha X_n(w) \geq 0, \quad \forall n \in \mathcal{N} \end{array}\right) - \left(\begin{array}{ll} \sup\limits_{y} & -\frac{y}{4} \\ \text{s.t.} & \left\| \left[\begin{array}{c} X_n(w) - \kappa \\ \left(\frac{v_n+\alpha X_n(w)-y}{2}\right) \end{array}\right] \right\|_2 \leq \frac{v_n+\alpha X_n(w)+y}{2} \\ & \forall n \in \mathcal{N} \\ & y \geq 0 \end{array}\right) \leq \beta,$$

where η, u, v, w, y are the variables. Since the inf appears on the left hand side of the inequality and the sup is preceded by a negative sign, both can be dropped (see Remark 2.2), and the resulting equivalent constraint system is:

$$\left\{\begin{array}{l} \eta + u\rho + u\sum\limits_{n\in\mathcal{N}} q_n\varphi^*\left(\frac{v_n-\eta}{u}\right) + \frac{y}{4} \leq \beta \\ \left\| \left[\begin{array}{c} X_n(w) - \kappa \\ \left(\frac{v_n+\alpha X_n(w)-y}{2}\right) \end{array}\right] \right\|_2 \leq \frac{v_n+\alpha X_n(w)+y}{2}, \quad \forall n \in \mathcal{N} \\ v_n + \alpha X_n(w) \geq 0, \quad \forall n \in \mathcal{N} \\ u, y \geq 0. \end{array}\right.$$

Since the $\varphi^*(\cdot)$ functions are convex, the resulting system of constraints is a system of convex, second-order conic, and linear constraints. Correspondingly, the complexity of the combination of the standard deviation and the φ-divergence set in Table 2.1 is denoted as CP.

2.7.2 Portfolio management

We consider as first numerical application of our methodology a stylized portfolio optimization problem. In this problem, the aim is to maximize the (worst-case) mean return subject to a maximum risk measure level, in both a nominal and robust setting. We choose the risk measure to be the EVaR for its importance as an upper bound on both the VaR and the CVaR. Additionally, the use of EVaR allows us to illustrate the power of our approach to tackle two-layer uncertainty sets.

2.7.2.1 Formulation and derivations of the robust counterparts

There are M available assets and N joint return scenarios for these assets, where Y_i^n denotes the gross return on the i-th asset in the n-th scenario. The decision vector $w \in \mathcal{W} = \left\{w \in \mathbb{R}^M : \quad 1^T w = 1, \quad w \geq 0\right\}$ consists of the portfolio weights of assets where we assume that shortselling is not allowed. The portfolio return in the n-th scenario is $X_n(w) = \sum_{i=1}^M w_i Y_i^n$. The maximum (robust) EVaR level is z. The nominal optimization problem is then:

$$\begin{aligned} \max \quad & \mu \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} q_n \left(-X_n(w)\right) \leq -\mu \\ & \sup_{\tilde{p}\in\mathcal{P}_q} \sum_{n\in\mathcal{N}} \tilde{p}_n \left(-X_n(w)\right) \leq z \\ & w \in \mathcal{W}, \end{aligned} \tag{2.23}$$

where $\mathcal{P}_q$ is defined in the row of Table 2.2 corresponding to the EVaR. Problem (2.23) includes a constraint involving a sup term, which requires a reformulation to a tractable form. In the terminology of this paper, this constraint is equivalent to a robust constraint on the negative mean return with uncertainty set $\mathcal{P}_q$ defined by the Kullback-Leibler divergence, and can be reformulated using the results of Sections 2.5 and 2.6.

We proceed to the more difficult and, hence, more illustrative robust problem. It shows the unifying power of our approach, including the derivation of the support function of the combined uncertainty set. Moreover, the constraint on the worst-case portfolio return in the robust problem is of the same type as the constraint on the risk measure in the nominal problem and, thus, the corresponding reformulation is also similar.

The uncertainty set for the nominal probability distribution q is defined as the Pearson set around a vector r (see Table 2.5):

$$\mathcal{Q} = \left\{q \geq 0 : \quad 1^T q = 1, \quad \sum_{n\in\mathcal{N}} \frac{(q_n - r_n)^2}{r_n} \leq \rho_{\mathcal{Q}}\right\}.$$

This formulation satisfies the conditions for the set $\mathcal{Q}$ in Table 2.3 for the combined uncertainty set since all the defining constraints can be formulated as constraints on convex functions in q. The portfolio optimization problem is then:

$$\begin{aligned} \max \quad & \mu \\ \text{s.t.} \quad & \sum_{n\in\mathcal{N}} q_n \left(-X_n(w)\right) \leq -\mu, \quad \forall q \in \mathcal{Q} && \text{(2.24a)} \\ & \sup_{\tilde{p}\in\mathcal{P}_q} \sum_{n\in\mathcal{N}} \tilde{p}_n \left(-X_n(w)\right) \leq z, \quad \forall q \in \mathcal{Q} && \text{(2.24b)} \\ & w \in \mathcal{W}. \end{aligned}$$

We shall reformulate the two constraints in problem (2.24) to their tractable forms using the results of Sections 2.5 and 2.6.

Constraint (2.24a). This is a robust constraint on the negative mean return with uncertainty set $\mathcal{Q}$ being the Pearson set. The corresponding conjugate function (page 35) is given by:

$$f_*(v,w) = \begin{cases} 0 & \text{if } -X_n(w) \le v_n^1, \quad \forall n \in \mathcal{N} \\ -\infty & \text{otherwise.} \end{cases}$$

To obtain the support function of the Pearson set, we combine the result on the support functions of the φ-divergence sets (2.15) (page 37) with the definition of the modified χ^2 distance in Table 2.5 (page 57):

$$\delta^*\left(v^1 \,\middle|\, \mathcal{P}_q^\varphi\right) = \inf_{u_1 \ge 0, \eta} \eta + u_1 \left(\rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max\left\{ -1, \frac{v_n^1 - \eta}{u_1} + \frac{1}{4}\left(\frac{v_n^1 - \eta}{u_1}\right)^2 \right\} \right).$$

Inserting the results on the conjugate and the support into (2.9) yields the tractable robust counterpart of (2.24a):

$$\begin{cases} \eta + u_1 \rho_{\mathcal{Q}} + \sum\limits_{n \in \mathcal{N}} r_n \max\left\{ -u_1, v_n^1 - \eta + \frac{1}{4}\frac{\left(v_n^1 - \eta\right)^2}{u_1} \right\} \le -\mu \\ u_1 \ge 0 \\ -X_n(w) \le v_n^1, \quad \forall n \in \mathcal{N}, \end{cases}$$

where η, u_1, v^1, w are the variables.

Constraint (2.24b). This is a robust constraint on the EVaR with $\mathcal{Q}$ defined as the Pearson set. We shall use the results for the EVaR (page 35) and the combined uncertainty set (page 39). The conjugate function $f_*(v,w)$ is thus the same as in the case of (2.24a). For the support function of the combined uncertainty set, we insert into the formula (2.18) for the support function of a combined uncertainty set the following components:

i conjugates related to the condition that the components of q should sum up to 1, that is, convex conjugates of $h_1(q) = 1^T q - 1$ and $h_2(q) = 1 - 1^T q$,

ii the convex conjugate of the modified χ^2 distance (Table 2.5, page 57), that is, $h_3^*(s) = \max\{-1, s + s^2/4\}$.

As a result, $\delta^*\left(\left(A^{\mathrm{C}}\right)^T v \middle| \mathcal{U}^{\mathrm{C}}\right)$ is equal to:

$$
\begin{aligned}
\inf_{\substack{\{u_i, v^i\} \\ i=2,\ldots,7}} \quad & u_2 - u_3 - u_4 \log \alpha + u_5 - u_6 + u_7 \left(\rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max\left\{-1, \tfrac{v^7_{N+n}}{u_7} + \tfrac{1}{4}\left(\tfrac{v^7_{N+n}}{u_7}\right)^2\right\}\right) \\
\text{s.t.} \quad & v^2_{1:N} \leq u_2 1 \\
& v^3_{1:N} \leq -u_3 1 \\
& v^5_{N+1:2N} \leq u_5 1 \\
& v^6_{N+1:2N} \leq -u_6 1 \\
& v^i_{N+1:2N} \leq 0, \quad i = 2, 3 \\
& v^i_{1:N} \leq 0, \quad i = 5, 6, 7 \\
& v^4_{N+n} + u_4 \left(\exp\left(\tfrac{v^4_n}{u_4}\right) - 1\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=2}^{7} v^i = \left(A^{\mathrm{C}}\right)^T v \\
& u_i \geq 0, \quad i = 2, \ldots, 7.
\end{aligned}
$$

Inserting the results on the conjugate and the support function into (2.9) yields the tractable robust counterpart of (2.24b):

$$
\begin{cases}
u_2 - u_3 - u_4 \log \alpha + u_5 - u_6 + u_7 \rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max\left\{-u_7, v^7_{N+n} + \frac{1}{4}\frac{\left(v^7_{N+n}\right)^2}{u_7}\right\} \leq z \\
v^2_{1:N} \leq u_2 1 \\
v^3_{1:N} \leq -u_3 1 \\
v^5_{N+1:2N} \leq u_5 1 \\
v^6_{N+1:2N} \leq -u_6 1 \\
v^i_{N+1:2N} \leq 0, \quad i = 2, 3 \\
v^i_{1:N} \leq 0, \quad i = 5, 6, 7 \\
v^4_{N+n} + u_4 \left(\exp\left(\tfrac{v^4_n}{u_4}\right) - 1\right) \leq 0, \quad \forall n \in \mathcal{N} \\
\sum_{i=2}^{7} v^i = \left(A^{\mathrm{C}}\right)^T v \\
u_i \geq 0, \quad i = 2, \ldots, 7 \\
-X_n(w) \leq v_n, \quad \forall n \in \mathcal{N},
\end{cases}
$$

with variables u_i, v^i, $i = 2, \ldots, 7$, v, w. We remark that it was possible to remove the inf term in the support function formulation due to its position on the left-hand side of the constraint. All the constraints in the above counterpart are convex in the decision variables, so in the terminology of Table 2.1 the complexity symbol of the system would be CP.

Combining the tractable robust counterparts of the constraints (2.24a) and (2.24b) with the rest of the problem formulation, we obtain that problem (2.24) is equivalent

to:

$$
\begin{aligned}
\max_{\substack{v,u_i,v^i,i=1,\ldots,7\\ w,\eta,\mu}} \quad & \mu \\
\text{s.t.} \quad & \eta + u_1\rho_{\mathcal{Q}} + \sum_{n\in\mathcal{N}} r_n \max\left\{-u_1, v_n^1 - \eta + \tfrac{1}{4}\tfrac{\left(v_n^1-\eta\right)^2}{u_1}\right\} \leq -\mu \\
& u_2 - u_3 - u_4 \log\alpha + u_5 - u_6 + u_7\rho_{\mathcal{Q}} + \\
& \qquad + \sum_{n\in\mathcal{N}} r_n \max\left\{-u_7, v_{N+n}^7 + \tfrac{1}{4}\tfrac{\left(v_{N+n}^7\right)^2}{u_7}\right\} \leq z \\
& v_{1:N}^2 \leq u_2 1 \\
& v_{1:N}^3 \leq -u_3 1 \\
& v_{N+1:2N}^5 \leq u_5 1 \\
& v_{N+1:2N}^6 \leq -u_6 1 \\
& v_{N+1:2N}^i \leq 0, \quad i = 2,3 \\
& v_{1:N}^i \leq 0, \quad i = 5,6,7 \\
& v_{N+n}^4 + u_4\left(\exp\left(\tfrac{v_n^4}{u_4}\right) - 1\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=2}^{7} v^i = \left(A^{\mathrm{C}}\right)^T v \\
& -X_n(w) \leq v_n, \quad \forall n \in \mathcal{N} \\
& -X_n(w) \leq v_n^1, \quad \forall n \in \mathcal{N} \\
& u_i \geq 0, \quad i = 1,\ldots,7 \\
& w \in \mathcal{W}.
\end{aligned}
\tag{2.25}
$$

This problem involves linear, convex quadratic, and convex constraints in the decision variables.

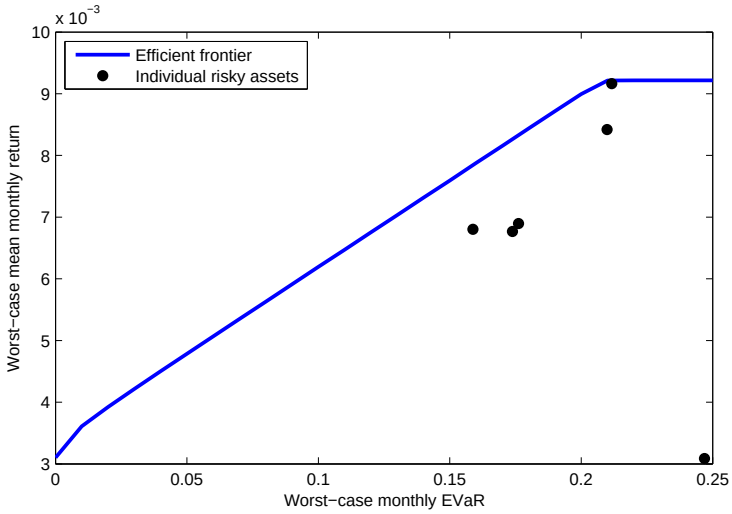
2.7.2.2 Numerical illustration

As a numerical illustration, we use 6 risky assets and 1 riskless asset, with data obtained from the website of Kenneth M. French.[1] The monthly data consists of 360 observations from February 1984 to January 2014.

The nominal distribution of the return scenarios assigns probability $r_n = \frac{1}{360}$ to each of the scenarios. We take $\alpha = 0.05$, which makes the EVaR an upper bound for the VaR and CVaR at level 0.05. The degree of uncertainty about the distribution of q in the robust model is defined by $\rho_{\mathcal{Q}} = 0.005$. This value has been chosen for illustration purposes - for high values of $\rho_{\mathcal{Q}}$ the best worst-case return of portfolios involving substantial fractions of the risk assets is negative. For that reason, for higher $\rho_{\mathcal{Q}}$, constraints on the EVaR will not be active as the optimal robust portfolio

[1] Available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The six risky assets are the '6 Portfolios Formed on Size and Book-to-Market (2 x 3)'. See the file with this name for a detailed description. As the riskless asset, we use the one-month US Treasury bill rate.

Figure 2.1 – Worst-case EVaR - worst-case mean return efficient frontier in our experiment. The dots illustrate the worst-case EVaR/mean returns of individual risky assets.



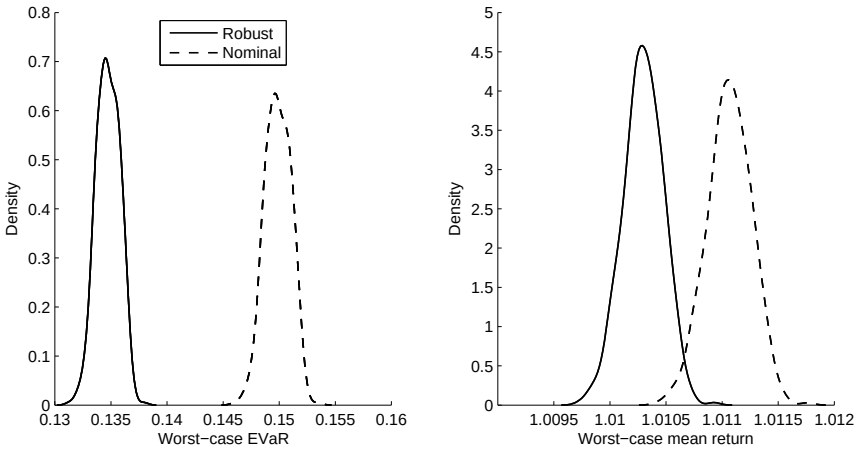
will consist of mostly the riskless asset in order to obtain a nonnegative worst-case return. Since the dataset includes two major crises (the dot.com crisis in 2000 and the financial crisis of 2007/2008), this indicates that on datasets that include such periods, portfolios whose key performance measure is the mean return, can be very conservative in a situation of substantial distributional uncertainty. Note also that in the case of the constraint on the EVaR, the size of the 'true' uncertainty set is much larger as it has a two layer combined set structure - the first layer defined with $\rho_{\mathcal{Q}} = 0.005$ and the second one with $\rho = -\log(0.05) \approx 3$.

A potential drawback of this method (historical simulation) is the use a discretized set of portfolio returns. The actual outcomes of portfolio returns typically will not coincide with any of the observations used in portfolio construction. Nevertheless, some authors show that such an approach might yield good results. For instance, Hanasusanto and Kuhn [28] use a distributional uncertainty set modelled with a χ^2 distance in a data-driven dynamic programming setting and show that the decisions obtained under the assumption of distributional uncertainty of the data-driven sample of the uncertain parameter value under consideration, exhibit substantial stability with respect to the sample used.

We solve problem (2.25) for values of $z = 0, 0.01, \ldots, 0.25$. In this way, we obtain the worst-case EVaR - worst-case mean return frontier. In addition to that, we compute the worst-case EVaR and worst-case mean return of the risky asset. Figure 2.1 presents both the frontier and the points corresponding to the six individual risky assets.

As it turns out, there is a single asset that dominates all other five risky assets, and that lies on the efficient frontier. An implication of this is that for all values of z

Figure 2.2 – Kernel density plots of the mean return and EVaR value of the nominal and robust portfolios obtained for $z = 0.15$.



the optimal portfolio consists of a mixture of the riskfree asset and this 'best' risky asset. In such a setting, the portfolio optimizer has no incentive to diversify the risk among the six risky assets.

To compare the performance of robust and nominal portfolios under distributional uncertainty, we conduct the following bootstrap experiment. We take the nominal (obtained by solving problem (2.23)) and robust portfolios for the maximum EVaR value $z = 0.15$. Then, we sample 500 probability distributions q around the nominal distribution r as follows: for $n = 1, ..., N-1$ the value r_n is sampled from a normal distribution with mean $r_n = \frac{1}{360}$ and standard deviation $\sqrt{\frac{\rho_{\mathcal{Q}}}{N^2}}$ and the last element is set $q_N = 1 - \sum_{n=1}^{N-1} q_n$. If it holds that $q \geq 0$, then the given vector is accepted. Out of this sample, 85% belonged to $\mathcal{Q}$. For each such q, we compute the EVaR and the mean return on the nominal and the robust portfolios. Figure 2.2 shows the results of the experiment.

The portfolios show significant differences in the distribution of their return and the EVaR value. In the left panel, the nominal portfolio violates the 0.15 upper bound in a large number of cases, whereas the robust portfolio's EVaR values oscillate in a region relatively far from 0.15. At the same time, the robust portfolio does not reveal any overconservatism - it is possible to find such q and $\tilde{p}$ that the EVaR of the robust portfolio is equal to 0.15. In the right panel we can see that on average the nominal portfolio has a significantly higher mean return. The differences between the means of EVaR and the return distributions are statistically significant at the 99% level.

2.7.3 Data-driven antenna array design

2.7.3.1 CVaR and variation distance

As next application, we consider the use of CVaR to approximate chance constraints in a distributionally uncertain antenna design problem. For that, we need the tractable robust counterpart of the constraint on the CVaR with uncertainty set $\mathcal{P}$ defined with the variation distance (one of the φ-divergence measures), see Table 2.5 in Appendix 2.C.1. Such a constraint is given by:

$$\inf_{\kappa} -\kappa + \frac{1}{\alpha}\sum_{n=1}^{N} p_n \max\left\{0, \kappa - X_i(w)\right\} \leq \beta, \quad \forall p \in \mathcal{P}$$

where

$$\mathcal{P} = \left\{ p: \quad p \geq 0, \quad 1^T p = 1, \quad \sum_{n\in\mathcal{N}} |p_n - q_n| \leq \rho \right\}. \tag{2.26}$$

To construct the robust counterpart we use: (1) the expression for the support function of $\mathcal{P}$, given in Section 2.6, page 37, with the relevant convex conjugate provided in Table 2.5 in Appendix 2.C.1; (2) the expression for the concave conjugate of the risk measure, discussed as a special case of the Optimized Certainty Equivalent in Section 2.5, page 32. Then, the robust counterpart is given by

$$\begin{cases} -\kappa + \eta + u\rho + \sum\limits_{n\in\mathcal{N}} q_n \max\{-u, v_n - \eta\} \leq \beta \\ v_n - \eta \leq u, \quad \forall n \in \mathcal{N} \\ v_n \geq \frac{1}{\alpha}\max\{0, \kappa - X_n(w)\}, \quad \forall n \in \mathcal{N} \\ u \geq 0. \end{cases} \tag{2.27}$$

2.7.3.2 Antenna design

In this section we consider an antenna design problem, adopted from [6]. The setting of the problem is as follows. There are N_A ring-shaped antennas belonging to the XY plane in $\mathbb{R}^3$. The radius of the k-th antenna is defined as k/N_A and the diagram $D(\varphi)$ of the antenna array is defined as a sum of diagrams $D_k(\varphi)$ of the antennas, with $D_k(\varphi)$ given by:

$$D_k(\varphi) = \frac{1}{2}\int_0^{2\pi} \cos\left(\frac{2\pi k}{N_A}\cos(\varphi)\cos(\gamma)\right) d\gamma.$$

The objective of the problem is to minimize the maximum of the diagram modulus in the angle of interest $0 \leq \varphi \leq 70°$:

$$\max_{0\leq\varphi\leq 70°} \left| \sum_{k=1}^{N_A} w_k D_k(\varphi) \right|,$$

subject to the restrictions that:

- the diagram in the interval $77^\circ \leq \varphi \leq 90^\circ$ is nearly uniform:

$$0.9 \leq \sum_{k=1}^{N_A} w_k D_k(\varphi) \leq 1, 77^\circ \leq \varphi \leq 90^\circ$$

- the diagram in other angles is not too large:

$$\left| \sum_{k=1}^{N_A} w_k D_k(\varphi) \right| \leq 1, \quad 70^\circ \leq \varphi \leq 77^\circ.$$

The problem is thus:

$$\begin{array}{ll} \min\limits_{\tau,w} & \tau \\ \text{s.t.} & -\tau \leq \sum\limits_{k=1}^{N_A} D_k(\varphi_i) w_k \leq \tau, \quad 0^\circ \leq \varphi_i \leq 70^\circ \\ & -1 \leq \sum\limits_{k=1}^{N_A} D_k(\varphi_i) w_k \leq 1, \quad 70^\circ \leq \varphi_i \leq 77^\circ \\ & 0.9 \leq \sum\limits_{k=1}^{N_A} D_k(\varphi_i) w_k \leq 1, \quad 77^\circ \leq \varphi_i \leq 90^\circ, \end{array} \tag{2.28}$$

$i = 1, \ldots, N_G$, where $\varphi_1, ..., \varphi_{N_G}$ is a 'fine grid' of equidistance placed points on $[0^\circ, 90^\circ]$.

However, we assume that a multiplicative implementation error affects the decision variable related to the k-th antenna:

$$w_k \mapsto \tilde{w}_k = (1 + z_k) w_k, \quad k = 1, \ldots, N_A.$$

The implementation error z_k consists of two parts: (1) a general error ζ affecting all antennas with the same power; (2) an idiosyncratic error δ_k, specific for each antenna:

$$z_k = \zeta + \delta_k,$$

where ζ and δ_k are independent, normally distributed with zero means and standard deviations σ_1 and σ_2 (σ_2 is the same for $k = 1, \ldots, N_A$), respectively. This is in line with the fact that in complex electrical systems, a part of the implementation error is common for all elements.

Assume there are N past observations of the errors, with the n-th sample denoted as $z^n = (\widehat{z}_1^{(n)}, \ldots, \widehat{z}_{N_A}^{(n)})$. Assume that z^n occurs with an uncertain probability, p_n, with nominal value $q_n = 1/N$ (each sample having equal probability), for each $n \in \mathcal{N}$. In such a setting, the random variables we consider are the diagrams of the antenna arrays at angles φ_i, as in the constraints in (2.28). For a given i, the n-th outcome of

the random antenna array diagram is equal to $X_n^i(w) = \sum_k (1 + \hat{z}_k^{(n)}) w_k D_k(\varphi_i)$, and its probability of occurrence is p_n.

We want each of the constraints in (2.28) to hold with at least probability $(1-\alpha)100\%$ over the sample at hand. For the first constraint in (2.28) this is equivalent to:

$$\sum_{n:X_n^i(w)\leq\tau} p_n \geq 1-\alpha, \quad \forall p \in \mathcal{P}.$$

This constraint is in fact equivalent to a constraint on the Value-at-Risk, as noted by [39] and it requires an integer linear programming reformulation. A tractable way to safely approximate chance constraints is the use of the CVaR. Then, the constraint is given by

$$\left\{ \mathrm{CVaR}_{p,\alpha}(-X^i(w)) \leq -\tau \quad \Leftrightarrow \quad \inf_{\kappa} -\kappa + \sum_{n=1}^{N} p_n \max\left\{0, \kappa - X_n^i(w)\right\} \leq -\tau \right\}, \quad \forall p \in \mathcal{P}.$$

We assume $\mathcal{P}$ to be specified as a variation distance set around q, as in (2.26). This choice is motivated by the fact that such $\mathcal{P}$ is linearly representable in p and q, and the resulting robust counterpart is a system of LP-representable constraints. Since in practice the past samples may be very large, computational tractability is one of the primary criteria for the set to choose. The other constraints are reformulated in a similar fashion. In this way, the problem to be solved becomes:

$$\begin{array}{lll}
\min\limits_{\tau,w} & \tau & \\
\text{s.t.} & \mathrm{CVaR}_{\alpha}(\tau - \sum\limits_k D_k(\varphi_i)\tilde{w}_k) \leq 0, & \forall p \in \mathcal{P}, \quad 0^\circ \leq \varphi_i \leq 70^\circ \\
& \mathrm{CVaR}_{p,\alpha}(\sum\limits_k D_k(\varphi_i)\tilde{w}_k + \tau) \leq 0, & \forall p \in \mathcal{P}, \quad 0^\circ \leq \varphi_i \leq 70^\circ \\
& \mathrm{CVaR}_{p,\alpha}(1 - \sum\limits_k D_k(\varphi_i)\tilde{w}_k) \leq 0, & \forall p \in \mathcal{P}, \quad 70^\circ \leq \varphi_i \leq 77^\circ \\
& \mathrm{CVaR}_{p,\alpha}(\sum\limits_k D_k(\varphi_i)\tilde{w}_k + 1) \leq 0, & \forall p \in \mathcal{P}, \quad 70^\circ \leq \varphi_i \leq 77^\circ \\
& \mathrm{CVaR}_{p,\alpha}(1 - \sum\limits_k D_k(\varphi_i)\tilde{w}_k) \leq 0, & \forall p \in \mathcal{P}, \quad 77^\circ \leq \varphi_i \leq 90^\circ \\
& \mathrm{CVaR}_{p,\alpha}(\sum\limits_k D_k(\varphi_i)\tilde{w}_k - 0.9) \leq 0, & \forall p \in \mathcal{P}, \quad 77^\circ \leq \varphi_i \leq 90^\circ,
\end{array} \tag{2.29}$$

where $i = 1, \ldots, N_G$. We consider $N_A = 40$ antennas and a sample of past $N = 200$ error vectors sampled with $\sigma_1 = 0.005$ and $\sigma_2 = 0.0025$, which implies dominance of the common error over the idiosyncratic error. In such a setting, we solve the distributionally robust problem (2.29) with $\alpha = 0.1$, and uncertainty levels $\rho \in \{0, 0.01, \ldots, 0.1\}$. The value $\rho = 0$ corresponds to the problem with no distributional uncertainty. Higher values of ρ would not change the solution as already $\rho = 0.1$ implies that the constraint on CVaR with $\alpha = 0.1$ is in fact a constraint that is to hold for every sampled $z^{(n)}$.

We investigate two questions:

1. On the given random sample of past errors, what is the impact of distributional uncertainty on the probability that at least one of the constraints is violated, for solutions assuming and not assuming distributional uncertainty?

2. How do solutions constructed with and without the assumption on distributional uncertainty perform out-of-sample, that is, with implementation error sampled from the original normal distribution?

To study the first question, for each of the solutions we conduct a simulation study where 10^5 error scenarios are bootstrapped from the sample in the following fashion. First, 100 probability distributions $\widehat{p}$ are sampled in such a way that $\widehat{p}_i$ is sampled from the normal distribution with mean q_i and standard deviation equal to $\rho/2$ for $i = 1, \ldots, N-1$, and p_N is defined as $\widehat{p}_N = 1 - \sum_{n=1}^{N-1} \widehat{p}_n$. If it holds that $\widehat{p} \geq 0$, the given vector is accepted as the probability distribution. For each probability distribution we sample subsequently 1000 error scenarios from the sample used to construct a given solution.

To study the second question, for each solution we sample 10^5 error vectors from the normal distribution as the sample drawn used to solve the problem. For both samples, we compute then the average probability of violating at least one problem's constraint.

Table 2.4 presents the results on the optimal value of the objective function and probabilities of violating at least one constraint. What can be observed is that the differences in the optimal value of the objective function are relatively small, ranging from 6.66 for $\rho = 0$ to 6.85 for $\rho = 0.1$. At the same time the robust solutions exhibit much smaller probabilities of at least one constraint being violated. For example, for the in-sample bootstrap the difference between the nominal solution and the robust solution with $\rho = 0.1$ is 35.88% compared to 23.92%.

What is even more interesting is that the robust solutions perform also consistently better than the nominal solution on out-of-sample implementation errors, with the biggest difference being 34.70% compared to 42.73%. Comparing the first ($\rho = 0$) and the last solution ($\rho = 0.1$) we see that the worst-case objective value which is 2.85% worse provides probability guarantees better by 33% on the in-sample errors and by 17% on the out-of-sample errors.

2.8 Conclusions

Constraints on risk measures of decision-dependent random variables under distributional uncertainty arise in numerous fields, such as economics, finance, and engineering. In this paper we have reviewed the literature on the problem of reformulating such constraints into tractable forms. As our contribution, we have provided a unified

Table 2.4 – Results of the antenna design experiment. 'In-sample bootstrap' denotes the probability of violating at least one constraint when the implementation error is sampled from the sample used for optimization. Similarly, 'Out-of-sample bootstrap' denotes the probability of violating at least one constraint when the implementation error is sampled from the original normal distribution for the implementation error.

		Measure	
ρ	Objective value ($\cdot 10^{-2}$)	In-sample bootstrap violation probability (%)	Out-of-sample bootstrap violation probability (%)
0 (no uncertainty)	6.66	35.88	42.73
0.01	6.66	29.82	41.19
0.02	6.70	28.89	39.49
0.03	6.73	30.90	39.09
0.04	6.76	26.91	37.84
0.05	6.78	27.85	35.95
0.06	6.80	24.87	34.70
0.07	6.81	27.84	35.04
0.08	6.83	26.91	34.97
0.09	6.84	23.96	35.26
0.1	6.85	23.92	35.47

framework for tackling this issue, showing that for many risk measures and statistically based uncertainty sets the constraints' components corresponding to the risk measure and to the uncertainty set can be separated. We have also demonstrated that this framework can be applied to risk measures that are nonlinear in the probability vector. In this way, for risk measures and uncertainty sets for which we provide a closed-form tractable robust counterpart and its complexity, our framework covers the results obtained up to now in the literature (see Table 2.1).

To provide the decision maker with a clear overview of available techniques, we summarize the complexity results obtained with our framework in combination with results already obtained in the literature in Table 2.1. These results can provide a useful guideline for researchers and practitioners of various backgrounds.

There are two issues that we find of particular importance when applying robust optimization to risk measures. Following the work of Wozabal [51], who analyzes the Wasserstein distance, it is interesting to investigate whether our framework can be extended to the case with continuous probability distributions, without converting continuous probability distributions into discrete ones.

Second, for the risk measures that we have not been able to analyze successfully one could investigate their sensitivity to the uncertainty considered in this paper. It may turn out that these risk measures themselves are sufficiently robust or that different tools are needed to develop computationally tractable robust constraints in terms of these risk measures.

Appendices

2.A Fenchel duality

Assume $f : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$ is a closed concave function and $g : \mathbb{R}^N \mapsto \mathbb{R} \cup \{+\infty\}$ is a closed convex function, with $f_*(\cdot)$ and $g^*(\cdot)$ being their concave and convex conjugates, respectively. Define the following primal problem

$$\sup \{f(x) - g(x) \,|x \in \mathrm{dom}(f) \cap \mathrm{dom}(g)\} \tag{P}$$

and its dual problem

$$\inf \{g^*(x) - f_*(x) \,|x \in \mathrm{dom}(f_*) \cap \mathrm{dom}(g^*)\}. \tag{D}$$

Then, the following theorem holds:

Theorem 2.2 *If $ri(dom(f)) \cap ri(dom(g)) \neq \emptyset$, then the optimal values of (P) and (D) are equal and the minimal value of (D) is attained.*
If $ri(dom(f_*)) \cap ri(dom(g^*)) \neq \emptyset$, then the optimal values of (P) and (D) are equal and the maximal value of (P) is attained.

2.B Conjugates of the risk measures

2.B.1 Necessary lemmas

First result presented here is taken from [41] (see his Corollary 37.3.2). It allows us to interchange the inf and sup terms in the worst-case formulations of the Optimized Certainty Equivalent, mean absolute deviation from the median, variance less the mean, and standard deviation less the mean.

Lemma 2.2 *[41, Corollary 37.3.2] Let C and D be nonempty closed convex sets in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively and let K be a continuous finite concave-convex function on $C \times D$. Then, if either C or D is bounded, one has:*

$$\inf_{v \in D} \sup_{u \in C} K(u, v) = \sup_{u \in C} \inf_{v \in D} K(u, v).$$

For the derivation of the conjugate function of the standard deviation less the mean we also need the following results.

Lemma 2.3 *Assume that $f_i(\cdot), i = 1, \ldots, m$, are concave, and the intersection of the relative interiors of the effective domains of $f_i(\cdot), i = 1, \ldots, m$ is nonempty, i.e., $\cap_{i=1}^m ri(domf_i) \neq \emptyset$. Then,*

$$\left(\sum_{i=1}^m f_i\right)_*(v) = \sup_{v_i, i=1,\ldots,m} \left\{ \sum_{i=1}^m (f_i)_*(v^i) \,\middle|\, \sum_{i=1}^m v^i = v \right\}.$$

Lemma 2.4 *[41, Theorem 16.3] Let B be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ be a concave function. Assume there exists an x such that $Bx \in ri(domf)$. Then, it holds that:*

$$(fB)_*(z) = \sup_y \left\{ f_*(y) \,\middle|\, B^T y = z \right\},$$

where for each z the supremum is attained, and where the function fB is defined by $(fB)(x) = f(Bx)$.

2.B.2 Standard deviation less the mean

In the case of the standard deviation less the mean we study the function:

$$f(p, w) = \sqrt{\sum_{n\in\mathcal{N}} p_n \left(X_n(w) - \kappa\right)^2} - \alpha \sum_{n\in\mathcal{N}} p_n X_n(w).$$

As an exception, we define the effective domain of $f(p, w)$ to be:

$$\mathrm{dom} f = \left\{ (p, w) : \quad \sum_{n\in\mathcal{N}} p_n \left(X_n(w) - \kappa\right)^2 \geq 0, \quad w \in \mathbb{R}^M \right\}.$$

For this particular case it is easier to operate in this setting and the results are still valid in combination with any uncertainty sets. We use Lemmas 2.3 and 2.4 with

$$f(p, w) = f_1(p, w) + f_2(p, w), \quad f_1(p, w) = -\alpha \sum_{n\in\mathcal{N}} p_n X_n(w), \quad f_2(p, w) = \sqrt{b^T p},$$

where

$$b = \left[\left(X_1(w) - \kappa\right)^2, \ldots, \left(X_N(w) - \kappa\right)^2 \right]^T.$$

We have that

$$\begin{aligned} (f_1)_*(v^1, w) &= \begin{cases} 0 & \text{for } v_n^1 = -\alpha X_n(w), \quad n \in \mathcal{N} \\ -\infty & \text{otherwise} \end{cases} \\ (f_2)_*(v^2, w) &= \sup \left\{ -\frac{1}{4s} \,\middle|\, bs = v^2, s \geq 0 \right\}. \end{aligned}$$

Then, substituting $v^1 = u$ and $s = 1/y$, by Lemma 2.3 we obtain:

$$\begin{aligned} f_*(v, w) = \sup_y \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \tfrac{1}{y} \begin{bmatrix} \left(X_1(w) - \kappa\right)^2 \\ \vdots \\ \left(X_N(w) - \kappa\right)^2 \end{bmatrix} + u = v \\ & u_n = -\alpha X_n(w), \quad n \in \mathcal{N} \\ & y \geq 0. \end{aligned}$$

Since the objective in the above formulation is decreasing in y and the left-hand side of the first constraint is increasing in y, we can change the '=' sign in the first constraint to '$\leq$' and arrive at the following, equivalent formulation:

$$\begin{aligned} f_*(v,w) = \sup_y \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \tfrac{1}{y}\begin{bmatrix} (X_1(w)-\kappa)^2 \\ \vdots \\ (X_N(w)-\kappa)^2 \end{bmatrix} \leq v-u \\ & u_n = -\alpha X_n(w), \quad n \in \mathcal{N} \\ & y \geq 0. \end{aligned}$$

The first constraint can be reformulated using the result of [35] on hyperbolic constraints to obtain the following:

$$\begin{aligned} f_*(v,w) = \sup_y \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \left\| \begin{bmatrix} X_n(w)-\kappa \\ \left(\frac{v_n-u_n-y}{2}\right) \end{bmatrix} \right\|_2 \leq \tfrac{v_n-u_n+y}{2}, \quad n \in \mathcal{N} \\ & v_n - u_n \geq 0, \quad n \in \mathcal{N} \\ & u_n = -\alpha X_n(w), \quad n \in \mathcal{N} \\ & y \geq 0. \end{aligned}$$

To obtain the final result (2.14) in the main text, the equality constraints are eliminated by inserting the equalities involving u_n into other expressions. This result is also obtained in Example 28 in [10].

2.C Support functions of the uncertainty sets

2.C.1 Examples of φ-divergence functions

One of the types of uncertainty sets for the probabilities is defined using so-called φ-divergence functions. For the statistical background behind this tool we refer the reader to [9]. Table 2.5, adopted from [9], presents potential choices for the function $\varphi(\cdot)$ and its conjugate $\varphi^*(\cdot)$. Two specific cases are commonly known. These are: (1) the Kullback-Leibler divergence which defines an uncertainty set based on the likelihood ratio statistical test, (2) the modified χ^2-distance which defines an uncertainty set based on the χ^2 goodness of fit test, also known as the Pearson test.

Table 2.5 – Examples of φ-divergence functions and their convex conjugate functions. Table is taken from Ben-Tal et al. (2013).

Name	$\varphi(t), \quad t \geq 0$	$\varphi^*(s)$
Kullback-Leibler	$t \log t - t + 1$	$e^s - 1$
Burg entropy	$-\log t + t - 1$	$-\log(1-s), \quad s < 1$
χ^2 distance	$\frac{1}{t}(t-1)^2$	$2 - 2\sqrt{1-s}, \quad s < 1$
Modified χ^2 distance	$(t-1)^2$	$\begin{cases} -1 & s < -2 \\ s + s^2/4 & s \geq -2 \end{cases}$
Hellinger distance	$(\sqrt{t}-1)^2$	$\frac{s}{1-s}, \quad s < 1$
χ-divergence	$\lvert t-1\rvert^\theta$	$s + (\theta-1)\left(\frac{\lvert s\rvert}{\theta}\right)^{\theta/(\theta-1)}$
Variation distance	$\lvert t-1\rvert$	$\max\{-1, s\}, \quad s \leq 1$
Cressie-Read	$\frac{1-\theta+\theta t-t^\theta}{\theta(1-\theta)}, \quad t \neq 0, 1$	$\frac{1}{\theta}(1-s(1-\theta))^{\theta/(1-\theta)} - \frac{1}{\theta}, \quad s < \frac{1}{1-\theta}$

2.C.2 φ-divergence

For the φ-divergence function the uncertainty region is defined as

$$\mathcal{P}_q^\varphi = \{p : p \geq 0, \quad g_i(p) \leq 0, \quad i = 1, 2, 3\},$$

where

$$\begin{aligned} g_1(p) &= 1^T p - 1 \\ g_2(p) &= -1^T p + 1 \\ g_3(p) &= \sum_{n \in \mathcal{N}} q_n \varphi\left(\frac{p_n}{q_n}\right) - \rho. \end{aligned}$$

Now, the convex conjugates of these three functions over the domain $p \geq 0$ are needed.

We begin with the function $g_1(\cdot)$:

$$\begin{aligned} g_1^*(y) &= \sup_{p \geq 0} \left\{ y^T p - 1^T p + 1 \right\} \\ &= \sup_{p \geq 0} \left\{ (y-1)^T p + 1 \right\} \\ &= \begin{cases} 1 & \text{if } y - 1 \leq 0 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Analogously:

$$g_2^*(y) = \begin{cases} -1 & \text{if } y + 1 \leq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

For the third function the derivation is:

$$\begin{aligned} g_3^*(y) &= \sup_{p\geq 0}\left\{y^T p - \sum_{n\in\mathcal{N}} q_n\varphi\left(\tfrac{p_n}{q_n}\right)+\rho\right\} \\ &= \sup_{p\geq 0}\left\{\sum_{n\in\mathcal{N}} y_n p_n - q_n\varphi\left(\tfrac{p_n}{q_n}\right)\right\}+\rho \\ &= \rho+\sum_{n\in\mathcal{N}}\sup_{p_n\geq 0}\left\{y_n p_n - q_n\varphi\left(\tfrac{p_n}{q_n}\right)\right\} \\ &= \rho+\sum_{n\in\mathcal{N}}\sup_{t\geq 0} q_n\left\{y_n t-\varphi(t)\right\} \\ &= \rho+\sum_{n\in\mathcal{N}} q_n\varphi^*(y_n). \end{aligned}$$

Lemma 2.1 gives us:

$$\delta^*\left(v\,\middle|\,\mathcal{P}_q^\varphi\right)=\inf_{\{u_i,v^i\},i=1,2,3}\left\{u_1 g_1^*\left(\frac{v^1}{u_1}\right)+u_2 g_2^*\left(\frac{v^2}{u_2}\right)+u_3 g_3^*\left(\frac{v^3}{u_3}\right),\ \left|\ \sum_{i=1}^3 v^i=v,\ u_i\geq 0\right.\right\}. \tag{2.30}$$

Notice that by Lemma 2.1 we have

$$\begin{aligned} u_1 g_1^*\left(\frac{v^1}{u_1}\right) &= \begin{cases} u_1 & \text{for } v^1\leq u_1 1 \\ +\infty & \text{otherwise} \end{cases} \\ u_2 g_2^*\left(\frac{v^2}{u_2}\right) &= \begin{cases} -u_2 & \text{for } v^2\leq -u_2 1 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

From here, we get:

$$\begin{aligned} \delta^*\left(v\,\middle|\,\mathcal{P}_q^\varphi\right)=\inf_{\{u_i,v^i\},i=1,2,3}\ & u_1-u_2+u_3\left(\rho+\sum_{n\in\mathcal{N}} q_n\varphi^*\left(\tfrac{v_n^3}{u_3}\right)\right) \\ \text{s.t.}\ & v^1\leq u_1 1 \\ & v^2\leq -u_2 1 \\ & \sum_{i=1}^3 v^i = v \\ & u_i\geq 0,\quad i=1,2,3. \end{aligned} \tag{2.31}$$

The equality constraint can be eliminated by inserting $v_n^3=v_n-v_n^1-v_n^2$ for each $n\in\mathcal{N}$. We get:

$$\begin{aligned} \delta^*\left(v\,\middle|\,\mathcal{P}_q^\varphi\right)=\inf_{u_1,u_2,u_3,v^1,v^2}\ & u_1-u_2+u_3\left(\rho+\sum_{n\in\mathcal{N}} q_n\varphi^*\left(\tfrac{v_n-v_n^1-v_n^2}{u_3}\right)\right) \\ \text{s.t.}\ & v^1\leq u_1 1 \\ & v^2\leq -u_2 1 \\ & u_i\geq 0,\quad i=1,2,3. \end{aligned} \tag{2.32}$$

Since the functions $\varphi^*(\cdot)$ are nondecreasing, one can substitute $\eta = u_1 - u_2$ to obtain result (2.15) in the main text.

2.C.3 Kolmogorov-Smirnov

The relevant uncertainty set is:

$$\mathcal{P}_q^{\mathrm{KS}} = \left\{ p : p \geq 0, \quad 1^T p = 1, \quad \max_{n \in \mathcal{N}} \left| p^T 1^n - q^T 1^n \right| \leq \rho \right\}.$$

Since all the constraints in the definition of $\mathcal{P}_q^{\mathrm{KS}}$ are linear in p, the Kolmogorov-Smirnov set can be defined as:

$$\mathcal{P}_q^{\mathrm{KS}} = \{ p : \quad p \geq 0, \quad Dp \leq d \},$$

where $D \in \mathbb{R}^{(2N+2)\times N}$, $d \in \mathbb{R}^{2N+2}$ with:

$$\begin{array}{llll}
D_{1n} = 1, & d_1 = 1, & \forall n \in \mathcal{N} & \\
D_{2n} = -1, & d_2 = -1, & \forall n \in \mathcal{N} & \\
D_{2+n,i} = 1, & d_{2+n} = \rho + q^T 1^n, & \forall i \leq n, & n \in \mathcal{N} \\
D_{2+N+n,i} = -1, & d_{2+N+n} = \rho - q^T 1^n, & \forall i \leq n, & n \in \mathcal{N},
\end{array}$$

with the other components equal to 0. The support function is equal to:

$$\begin{aligned}
\delta^* \left(v \,\middle|\, \mathcal{P}^{\mathrm{KS}} \right) = \sup_p \quad & v^T p \\
\text{s.t.} \quad & Dp \leq d \\
& p \geq 0.
\end{aligned}$$

The final result (2.16) in the main text is obtained via strong LP duality.

2.C.4 Wasserstein

The definition of the Wasserstein set involves a variable matrix K, so that the set $\mathcal{U}$ is actually a set both in K and q. For that reason, we use an extended vector p' consisting of both these variables and 'extract' the vector p out of p' using a relevant A matrix. We take the extended vector to be:

$$p' = \left[p^T, K_1^T, K_2^T, ..., K_N^T \right]^T,$$

where $K_1, ..., K_N$ are the subsequent columns of K. A matrix A^{W} such that $A^{\mathrm{W}} p' = p$ is given by $A^{\mathrm{W}} = [I \,|0_{N \times N^2}]$. Since the constraints in the definition of $\mathcal{P}_q^{\mathrm{W}}$ are linear in (p, K), the Wasserstein set can be defined as:

$$\mathcal{U}_q^{\mathrm{W}} = \{ p' : \quad p' \geq 0, \quad Dp' \leq d \},$$

where $D \in \mathbb{R}^{(4N+3)\times N(N+1)}$, $d \in \mathbb{R}^{4N+3}$ and their entries are:

$$\begin{array}{lll}
D_{1n} = 1, & d_1 = 1, & \forall n \in \mathcal{N} \\
D_{2n} = -1, & d_2 = -1, & \forall n \in \mathcal{N} \\
D_{3,Ni+j} = \|Y_i - Y_j\|^d, & d_3 = \rho, & \forall i,j \in \mathcal{N} \\
D_{3+n,n} = -1, & D_{3+n,Nn+i} = 1, & \forall i,n \in \mathcal{N} \\
D_{3+N+n,n} = 1, & D_{3+N+n,Nn+i} = -1, & \forall i,n \in \mathcal{N} \\
D_{3+2N+n,Ni+n} = 1, & d_{3+2N+n} = -q_n, & \forall i,n \in \mathcal{N} \\
D_{3+3N+n,Ni+n} = -1, & d_{3+3N+n} = q_n, & \forall i,n \in \mathcal{N},
\end{array}$$

with the other components equal to 0. The support function is equal to:

$$\begin{aligned}
\delta^*\left(\left(A^{\mathrm{W}}\right)^T v \middle| \mathcal{U}_q^{\mathrm{W}}\right) = \sup_{p'} \quad & v^T A^{\mathrm{W}} p' \\
\text{s.t.} \quad & Dp' \leq d \\
& p' \geq 0.
\end{aligned}$$

From here, the final result (2.17) is obtained via strong LP duality.

2.C.5 Combined set

We substitute $p' = \left[p^T, q^T\right]^T$ so that $p = A^{\mathrm{C}} p'$, where $A^{\mathrm{C}} = [I|0_{N\times N}]$. The set $\mathcal{U}^{\mathrm{C}}$ is then:

$$\mathcal{U}^{\mathrm{C}} = \{p' : p' \geq 0, \quad g_i(p') \leq 0, \quad i = 1,2,3, \quad h_i(q) \leq 0, \quad i = 1,...,Q\}.$$

The first three convex functions from formulation of $\mathcal{U}^{\mathrm{C}}$ are:

$$\begin{aligned}
g_1(p') &= 1^T p - 1 \\
g_2(p') &= -1^T p + 1 \\
g_3(p') &= \sum_{n\in\mathcal{N}} q_n \varphi\left(\tfrac{p_n}{q_n}\right) - \rho.
\end{aligned}$$

The conjugates of the first two have been obtained for the φ-divergence set. Thus, only the third one remains:

$$\begin{aligned}
g_3^*(y) &= \sup_{p'\geq 0}\left\{y^T p' - g_3(p')\right\} \\
&= \sup_{p,q\geq 0}\left\{y_{1:N}^T p + y_{N+1:2N}^T q - \sum_{n\in\mathcal{N}} q_n \varphi\left(\tfrac{p_n}{q_n}\right) + \rho\right\} \\
&= \sup_{q\geq 0}\left\{y_{N+1:2N}^T q + \sup_{p\geq 0}\left\{y_{1:N}^T p - \sum_{n\in\mathcal{N}} q_n \varphi\left(\tfrac{p_n}{q_n}\right) + \rho\right\}\right\} \\
&= \sup_{q\geq 0}\left\{y_{N+1:2N}^T q + \sum_{n\in\mathcal{N}} q_n \sup_{u_n\geq 0}\left\{y_n u_n - \varphi\left(u_n\right)\right\} + \rho\right\} \\
&= \sup_{q\geq 0}\left\{\sum_{n\in\mathcal{N}} q_n\left(y_{N+n} + \varphi^*(y_n)\right) + \rho\right\} \\
&= \begin{cases} \rho & \text{for } y_{N+n} + \varphi^*(y_n) \leq 0 \quad \forall n \in \mathcal{N} \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Since all $h_i(\cdot)$ depend only on q, the support function of $\mathcal{U}^{\mathrm{C}}$ is given by (Lemma 2.1):

$$\begin{aligned}
\delta^*\left(\left(A^{\mathrm{C}}\right)^T v \,\middle|\, \mathcal{U}^{\mathrm{C}}\right) = \inf \quad & u_1 - u_2 + u_3\rho + \sum_{i=1}^{Q} u_{i+3} h_i^*\left(\tfrac{v_{N+1:2N}^{i+3}}{u_{i+3}}\right) \\
\text{s.t.} \quad & v_{1:N}^1 \leq u_1 1 \\
& v_{1:N}^2 \leq -u_2 1 \\
& v_{N+1:2N}^i \leq 0, \quad i = 1,2 \\
& v_{1:N}^i \leq 0, \quad i = 4,...,Q+3 \\
& \tfrac{v_{N+n}^3}{u_3} + \varphi^*\left(\tfrac{v_n^3}{u_3}\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=1}^{Q+3} v^i = v \\
& u_i \geq 0, \quad i = 1,...,Q+3.
\end{aligned}$$

The only thing left is to remove nonconvexity from the constraint $\frac{v_{N+n}^3}{u_3} + \varphi^*\left(\frac{v_n^3}{u_3}\right) \leq 0$. One can do that by multiplying both sides by u_3 to obtain the final result.

2.C.6 Anderson-Darling

The relevant set formulation is (see Table 2.3):

$$\mathcal{P}_{\mathrm{emp}}^{\mathrm{AD}} = \left\{p : p \geq 0, \quad g_i(p) \leq 0, \quad i = 1,2,3\right\},$$

where

$$\begin{aligned}
g_1(p) &= 1^T p - 1 \\
g_2(p) &= -1^T p + 1 \\
g_3(p) &= -N - \sum_{n\in\mathcal{N}} \tfrac{2n-1}{N}\left[\log\left(p^T 1^n\right) + \log\left(p^T 1^{-n}\right)\right] - \rho.
\end{aligned}$$

It is only necessary to derive the conjugate of $g_3(\cdot)$. Let us write $g_3(\cdot)$ as:

$$g_3(p) = \sum_{n\in\mathcal{N}} \left[-\left[\frac{2n-1}{N} \log\left(p^T 1^n\right) + \frac{\rho+N}{2N} \right] - \left[\frac{2n-1}{N} \log\left(p^T 1^{-n}\right) + \frac{\rho+N}{2N} \right] \right].$$

By results of [10], it is only needed to derive the convex conjugate of the function

$$H_n(t) = -\frac{2n-1}{N} \log(t) - \frac{\rho+N}{2N}, \quad t \geq 0.$$

It is given by:

$$\begin{aligned} H_n^*(s) &= \sup_{t\geq 0} \left\{ st + \tfrac{2n-1}{N} \log(t) + \tfrac{\rho+N}{N} \right\} \\ &= \begin{cases} -\tfrac{2n-1}{N} - \tfrac{2n-1}{N} \log\left(\tfrac{-Ns}{2n-1}\right) + \tfrac{\rho+N}{2N} & \text{if } s < 0 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Using Lemma 2.1, we obtain:

$$\begin{aligned} \delta^*\left(v \,\middle|\, \mathcal{P}_{\text{emp}}^{\text{AD}}\right) = \inf_{\substack{\{w^{n+},w^{n-}\},n\in\mathcal{N};\\ \{z_n^+,z_n^-\},n\in\mathcal{N};\\ u_1,u_2,u_3,v^1,v^2}} & -\sum_{n\in\mathcal{N}} \tfrac{(2n-1)u_3}{N} \left[2 + \log\left(\tfrac{-Nz_n^+}{(2n-1)u_3}\right) + \log\left(\tfrac{-Nz_n^-}{(2n-1)u_3}\right) \right] \\ & + u_3(\rho+N) + u_1 - u_2 \\ \text{s.t.} \quad & z_n^+ 1^n = w^{n+}, \quad \forall n \in \mathcal{N} \\ & z_n^- 1^{-n} = w^{n-}, \quad \forall n \in \mathcal{N} \\ & v^1 \leq u_1 1 \\ & v^2 \leq -u_2 1 \\ & \sum_{n\in\mathcal{N}} (w^{n+} + w^{n-}) + v^1 + v^2 = v \\ & z_n^+, z_n^- \leq 0, \quad \forall n \in \mathcal{N} \\ & u_1, u_2, u_3 \geq 0. \end{aligned}$$

We eliminate the equalities involving w_n^+ and w_n^- to obtain:

$$\begin{aligned} \inf_{\substack{\{w^{n+},w^{n-}\},n\in\mathcal{N},\\ u_1,u_2,u_3,v^1,v^2}} & -\sum_{n\in\mathcal{N}} \tfrac{(2n-1)u_3}{N} \left[2 + \log\left(\tfrac{-Nz_n^+}{(2n-1)u_3}\right) + \log\left(\tfrac{-Nz_n^-}{(2n-1)u_3}\right) \right] \\ & + u_3(\rho+N) + u_1 - u_2 \\ \text{s.t.} \quad & v^1 \leq u_1 1 \\ & v^2 \leq -u_2 1 \\ & \sum_{n\in\mathcal{N}} (z_n^+ 1^n + z_n^- 1^{-n}) + v^1 + v^2 = v \\ & z_n^+, z_n^- \leq 0, \quad \forall n \in \mathcal{N}. \\ & u_1, u_2, u_3 \geq 0. \end{aligned}$$

In the third constraint it is possible to change the equality into inequality because of the properties of the other constraints and the 'objective function'. Also, by the properties of the formulation above one can substitute $\eta = u_1 - u_2$ and remove the variables v^1, v^2. In this way result (2.19) in the main text is obtained.

2.C.7 Cramer-von Mises

The set definition is:

$$\mathcal{P}_{\text{emp}}^{\text{CvM}} = \left\{ p : p \geq 0, \quad 1^T p = 1, \quad \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left[\frac{2n-1}{2N} - p^T 1^n \right]^2 \leq \rho \right\},$$

which can be reformulated as $\mathcal{P}_{\text{emp}}^{\text{CvM}} = \{p : \quad g_i(p) \leq 0, \quad i = 1, ..., N+3\}$, where

$$\begin{aligned} g_1(p) &= 1^T p - 1 \\ g_2(p) &= -1^T p + 1 \\ g_3(p) &= p^T E p + b^T p + c \\ g_{3+n}(p) &= -p^T e^n, \quad \forall n \in \mathcal{N}, \end{aligned}$$

where

$$c = -\rho + \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} \right)^2, \qquad b = \begin{bmatrix} -\sum_{j=1}^{N} \frac{2j-1}{N} \\ -\sum_{j=2}^{N} \frac{2j-1}{N} \\ \vdots \\ -\sum_{j=N}^{N} \frac{2j-1}{N} \end{bmatrix},$$

and $E \in \mathbb{R}^{N \times N}$ is a positive definite matrix such that $E_{ij} = N + 1 - \max\{i, j\}$ for $i, j \in \mathcal{N}$.

Contrary to the previous cases, we assume the domains of the functions $g_i(\cdot)$ to be $\mathbb{R}^N$ and we include the nonnegativity constraints on p as explicit functional constraints $g_{3+n}(p) = -p^T e^n \leq 0$. Then, we derive the conjugates of $g_i(\cdot)$ as supremums over $p \in \mathbb{R}^N$, which makes the derivation of $g_3^*(\cdot)$ easier. The resulting formula for the support function is equivalent to the formula obtained using the standard assumption about the domains of $g_i(\cdot)$ which, however, would require much more algebra.

Remark 2.3 *Positive definiteness of E follows from the following transformations. Denote by $E^{(k)}$ a matrix for which*

$$E_{ij}^{(k)} = \begin{cases} 1 & \text{for } i, j \leq k \\ 0 & \text{otherwise.} \end{cases}$$

Consider $p \in \mathbb{R}^N$. We then have:

$$\begin{aligned} p^T E p &= p^T \left(\sum_{k=1}^{N} E^{(k)} \right) p \\ &= \sum_{k=1}^{N} p^T E^{(k)} p \\ &= \sum_{k=1}^{N} \left(p_1 + \cdots + p_k \right)^2 \geq 0, \end{aligned}$$

with 0 being attained if and only if $p_1 = \ldots = p_N = 0$.

It is important to note that the inverse of E has a tridiagonal structure, allowing for efficient computations.

We proceed to the derivations of the conjugates. These are:

$$\begin{aligned} g_3^*(y) &= \sup_p \left\{ y^T p - p^T E p - b^T p - c \right\} \\ &= \sup_p \left\{ -p^T E p - (b-y)^T p - c \right\} \\ &= \tfrac{1}{4}(b-y) E^{-1} (b-y) - c, \end{aligned}$$

and

$$\begin{aligned} g_{3+n}^*(y) &= \sup_p \left\{ y^T p + p^T e^n \right\} \\ &= \begin{cases} 0 & \text{if } y + e^n = 0 \\ +\infty & \text{otherwise} \end{cases} \end{aligned}$$

for all $n \in \mathcal{N}$. The support function is equal to:

$$\begin{aligned} \delta^* \left(v \,\middle|\, \mathcal{P}_{\text{emp}}^{\text{CvM}} \right) = \inf_{\{u_i, v^i\}, i=1,\ldots,N+3} \quad & u_1 - u_2 + \tfrac{1}{4} u_3 \left(b - \tfrac{v^3}{u_3} \right)^T E^{-1} \left(b - \tfrac{v^3}{u_3} \right) - u_3 c \\ \text{s.t.} \quad & v^1 = u_1 1 \\ & v^2 = -u_2 1 \\ & v^{3+n} = -u_{3+n} e^n, \quad n \in \mathcal{N} \\ & \sum_{i=1}^{N+3} v^i = v \\ & u_i \geq 0, \quad i = 1, ..., N+3. \end{aligned}$$

The 'objective function' in the above formulation, already convex in its arguments, can be transformed into a system of linear and second-order conic constraints. Indeed, one may introduce an extra variable $t \geq 0$ such that

$$u_3 \left(b - \frac{v^3}{u_3} \right)^T E^{-1} \left(b - \frac{v^3}{u_3} \right) \leq t \quad \Leftrightarrow \quad \frac{\left(u_3 b - v^3 \right)^T E^{-1} \left(u_3 b - v^3 \right)}{u_3} \leq t.$$

Then, introducing $z = u_3 b - v^3$ and $E^{-1} = P^T P$ (where P is a $N \times N$ matrix because of the positive definiteness of E) we obtain

$$\frac{\left(u_3 b - v^3\right)^T E^{-1} \left(u_3 b - v^3\right)}{u_3} \leq t \quad \Leftrightarrow \quad \frac{(Pz)^T (Pz)}{u_3} \leq t.$$

This can be transformed, using the results from [5], to:

$$\left\| \begin{bmatrix} Pz \\ \frac{t-u_3}{2} \end{bmatrix} \right\|_2 \leq \frac{t+u_3}{2}.$$

Implementing this and eliminating the equality constraints by inserting the equalities involving u_{3+n} into other places yields result (2.20) in the main text.

2.C.8 Watson

The set definition is:

$$\mathcal{P}_{\text{emp}}^{\text{Wa}} = \left\{ p : p \geq 0, 1^T p = 1, \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} - p^T 1^n \right)^2 - N \left(\frac{1}{N} \sum_{n \in \mathcal{N}} p^T 1^n - \frac{1}{2} \right)^2 \leq \rho \right\},$$

where the last constraint can be formulated as in the case of the Cramer-von Mises set, with parameter values:

$$c = -\rho + \tfrac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\tfrac{2n-1}{2N} \right)^2 - \tfrac{N}{4}, \qquad b = \begin{bmatrix} -\sum_{j=1}^{N} \frac{2j-1}{N} + N \\ -\sum_{j=2}^{N} \frac{2j-1}{N} + (N-1) \\ \vdots \\ -\sum_{j=N}^{N} \frac{2j-1}{N} + 1 \end{bmatrix},$$

and $E \in \mathbb{R}^{N \times N}$ such that $E_{i,j} = N + 1 - \max\{i, j\} - \frac{(N+1-i)(N+1-j)}{N}$ for all $i, j \in \mathcal{N}$. The matrix E is positive semidefinite with a one-dimensional nullspace, which we prove in the following remark.

Remark 2.4 *Assume $p \in \mathbb{R}^N$ and $d_n = p_1 + \ldots + p_n$ for $n \in \mathcal{N}$. We have:*

$$\begin{aligned} p^T E p &= \sum_{n=1}^{N} \left(p^T 1^n \right)^2 - \frac{1}{N} \left(\sum_{n=1}^{N} p^T 1^n \right)^2 \\ &= \sum_{n=1}^{N} d_n^2 - \frac{1}{N} \left(\sum_{n=1}^{N} d_n \right)^2 \\ &= N \left(\frac{\sum_{n=1}^{N} d_n^2}{N} - \left(\frac{\sum_{n=1}^{N} d_n}{N} \right)^2 \right) \geq 0, \end{aligned}$$

where the first equality follows from the definition of $\mathcal{P}^{Wa}_{emp}$ and the inequality follows from the inequality between arithmetic and quadratic means, and 0 is attained if and only if $d_1 = \ldots = d_N$, that is, when $p_2 = p_3 = \ldots = p_N$ with arbitrary p_1.

We proceed to the derivation of the support function $g^*_3(\cdot)$. It is:

$$\begin{aligned} g^*_3(y) &= \sup_p \left\{ y^T p - p^T E p - b^T p - c \right\} \\ &= \sup_p \left\{ -p^T E p - (b-y)^T p - c \right\} \\ &= \begin{cases} \frac{1}{4}(b-y)E^\dagger(b-y) - c & \text{if } (b-y) \in \mathrm{Im}E \\ +\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where $E^\dagger$ denotes a pseudo-inverse of E and ImE denotes the subspace spanned by the columns of E. From here on, the derivation is analogous to the case of the Cramer-von Mises test, with an extra constraint $(b-y) \in \mathrm{Im}E$, implemented as $\exists \lambda$ s.t. $b - y = E\lambda$.

2.C.9 Kuiper

The Kuiper set is defined by

$$\mathcal{P}^{\mathrm{K}}_{\mathrm{emp}} = \left\{ \max_{n \in \mathcal{N}} \left(\frac{n}{N} - p^T 1^n \right) + \max_{n \in \mathcal{N}} \left(p^T 1^{n-1} - \frac{n-1}{N} \right) \leq \rho \right\}.$$

Using additional variables z_1,z_2 it can be transformed to

$$\begin{aligned} \mathcal{U}^{\mathrm{K}}_{\mathrm{emp}} &= \Big\{ (p, z_1, z_2) : 1^T p = 1, \quad z_1 + z_2 \leq \rho, \\ &\quad \max_{n \in \mathcal{N}} \left(\tfrac{n}{N} - p^T 1^n \right) \leq z_1, \quad \max_{n \in \mathcal{N}} \left(p^T 1^{n-1} - \tfrac{n-1}{N} \right) \leq z_2 \Big\}. \end{aligned}$$

Thus, we use a vector $p' = \left[p^T, z_1, z_2 \right]^T$ and a matrix $A^{\mathrm{K}} = [I \,|0_{N \times 2}]$. The set $\mathcal{U}^{\mathrm{K}}_{\mathrm{emp}}$ is then:

$$\mathcal{U}^{\mathrm{K}}_{\mathrm{emp}} = \{ p' : p' \geq 0, \quad Dp' \leq d \},$$

where $D \in \mathbb{R}^{(2N+3)\times(N+2)}$, $d \in \mathbb{R}^{2N+3}$ are defined by:

$$\begin{array}{llll} D_{1,n} = 1, & d_1 = 1, & & \forall n \in \mathcal{N} \\ D_{2,n} = -1, & d_2 = -1, & & \forall n \in \mathcal{N} \\ D_{2+n,i} = -1, & D_{2+n,N+1} = -1, & d_{n+2} = -n/N, & \forall i \leq n, n \in \mathcal{N} \\ D_{N+2+n,i} = 1, & D_{N+2+n,N+2} = -1, & d_{N+2+n} = (n-1)/N, & \forall i \leq n-1, n \in \mathcal{N} \\ D_{2N+3,N+1} = 1, & D_{2N+3,N+2} = 1, & d_{2N+3} = \rho, & \end{array}$$

with all other components equal to 0. The final form (2.22) in the main text is obtained via strong LP duality.

Bibliography

[1] C. Acerbi. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, 2002.

[2] A. Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123, 2012.

[3] S. Ahmed, U. Cakmak, and A. Shapiro. Coherent risk measures in inventory problems. *European Journal of Operational Research*, 182(1):226–238, 2007.

[4] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[5] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

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

[7] A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: the optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.

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

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

[10] A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.

[11] D. Bertsimas and D. B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.

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

[13] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. http://arxiv.org/pdf/1401.0212.pdf, 2013.

[14] D. Bertsimas, V. Gupta, and N. Kallus. Robust SAA. http://arxiv.org/pdf/1408.4445v1.pdf, 2014.

[15] T. Breuer and I. Csiszár. Measuring distribution model risk. *Mathematical Finance*, 26(2):395–411, 2013.

[16] G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.

[17] G. C. Calafiore. A robust-CVaR optimization approach with application to breast cancer therapy. *European Journal of Operational Research*, 238(3):876–885, 2014.

[18] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[19] K. Dowd. *Measuring market risk*. John Wiley & Sons, 2007.

[20] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

[21] P. Embrechts, R. Frey, and A. McNeil. *Quantitative risk management*. Princeton Series in Finance, Princeton, 2005.

[22] L. G. Epstein. A definition of uncertainty aversion. *The Review of Economic Studies*, 66(3):579–608, 1999.

[23] A. Fertis, M. Baes, and H. J. Laethi. Robust risk management. *European Journal of Operational Research*, 222(3):663–672, 2012.

[24] H. Föllmer and A. Schied. Convex risk measures. In *Encyclopedia of Quantitative Finance*, pages 355–363. John Wiley & Sons, 2010.

[25] V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483, 2014.

[26] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

[27] J. Gotoh and S. Uryasev. Support vector machines based on convex risk functions and general norms. *Annals of Operations Research*, pages 1–28, 2016.

[28] G. A. Hanasusanto and D. Kuhn. Robust data-driven dynamic programming. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 827–835, 2013.

[29] Z. Hu and L. J. Hong. Kullback-Leibler divergence constrained distributionally robust optimization. http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf, 2013.

[30] Z. Hu, L. J. Hong, and A. M. C. So. Ambiguous probabilistic programs. http://www.optimization-online.org/DB_FILE/2013/09/4039.pdf, 2013.

[31] D. Huang, S. F. Zhu, F. J. Fabozzi, and M. Fukushima. Portfolio selection under distributional uncertainty: A relative robust CVaR approach. *European Journal of Operational Research*, 203(1):185–194, 2010.

[32] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Advance online publication in Mathematical Programming*, 2015.

[33] D. Klabjan, D. Simchi-Levi, and M. Song. Robust stochastic lot sizing by means of histograms. *Production and Operations Management*, 22(3):691–710, 2013.

[34] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models: from data to decisions.* John Wiley & Sons, 2012.

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

[36] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[37] K. Mosler and P. Bazovkin. Stochastic linear programming with a distortion risk constraint. *OR Spectrum*, 36(4):949–969, 2012.

[38] K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141, 2009.

[39] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[40] A. Pichler. Evaluations of risk measures for different probability measures. *SIAM Journal on Optimization*, 23(1):530–551, 2013.

[41] R. T. Rockafellar. *Convex analysis.* Princeton University Press, 1970.

[42] R. T. Rockafellar and J. O. Royset. Engineering decisions under risk averseness. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 1(2):04015003, 2015.

[43] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[44] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Risk tuning with generalized linear regression. *Mathematics of Operations Research*, 33(3):712–729, 2008.

[45] R. Rühlicke. *Robust risk management in the context of Solvency II Regulations.* PhD thesis, Universität Duisburg-Essen, 2013.

[46] A. Schied and H. Föllmer. *Stochastic finance: an introduction in discrete time.* Walter De Gruyter, 2004.

[47] W. F. Sharpe. Mutual fund performance. *Journal of Business*, 39:119–138, 1966.

[48] O. Thas. *Comparing distributions.* Springer, 2010.

[49] Z. Wang, P. Glynn, and Y. Ye. Likelihood robust optimization for data-driven problems. *Computational Management Science*, 2:241–261, 2016.

[50] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[51] D. Wozabal. Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research*, 62(6):1302–1315, 2014.

[52] S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.

CHAPTER 3

Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set

3.1 Introduction

Robust optimization (RO, see Ben-Tal et al. (2009)) has become one of the main approaches to optimization under uncertainty. One of its applications are multiperiod problems where, period after period, values of the uncertain parameters are revealed and new decisions are implemented. Adjustable Robust Optimization (ARO, see Ben-Tal et al. (2004)) addresses such problems by formulating the decision variables as functions of the revealed uncertain parameters. Ben-Tal et al. (2004) prove that without any functional restrictions on the form of adjustability, the resulting problem is NP-hard. For that reason, several functional forms of the decision rules have been proposed, with the most popular being the affinely adjustable decision rules. However, only for a limited class of problems do they yield problems that can be reformulated to a computationally tractable form (see Ben-Tal et al. (2009)). In particular, for problems without fixed recourse, where the later-period problem parameters depend also on the uncertain parameters from earlier periods, it is nontrivial to construct tractable decision rules. The difficulty level grows even more when the adjustable variables are binary or integer. Addressing this problem is the topic of our paper. We propose a simple and intuitive method to construct adjustable decision rules, applicable also to problems with integer adjustable variables and to problems without fixed recourse. For problems with fixed recourse our methodology can be combined with linear decision rules for the continuous decision variables.

The contribution of our paper is twofold. First, we propose a methodology of iterative splitting of the uncertainty set into subsets, for each of which a scalar later-period decision shall be determined. A given decision is implemented in the next period if the revealed uncertain parameter belongs to the corresponding subset. Using scalar decisions per subset ensures that the resulting problem has the same complexity as

the static robust problem. This approach provides an upper bound on the optimal value of the adjustable robust problem. Next to that, we propose a method of obtaining lower bounds, being a generalization of the approach of Hadjiyiannis et al. (2011).

As a second contribution, we provide theoretical results supporting the decision of how to split the uncertainty set into smaller subsets for problems with continuous decision variables. The theory identifies sets of scenarios for the uncertain parameters that have to be divided. On the basis of these results, we propose set-splitting heuristics for problems including also integer decision variables. As a side result, we prove the reverse of the result of Gorissen et al. (2014). Namely, we show that the optimal KKT vector of the tractable robust counterpart of a linear robust problem, obtained using the results of Ben-Tal et al. (2015), yields an optimal solution to the optimistic dual (see Beck and Ben-Tal (2009)) of the original problem.

ARO was developed to (approximately) solve problems with continuous variables. Ben-Tal et al. (2004) introduce the concept of using affinely adjustable decision rules and show how to apply such rules to obtain (approximate) optimal solutions to multiperiod problems. Affinely adjustable decisions turn out to be very effective for the inventory management example, which shall be also visible in the results of our paper. Their approach has been later extended to other function classes by Chen et al. (2007), Chen and Zhang (2009), Ben-Tal et al. (2009) and Bertsimas et al. (2011b). Bertsimas et al. (2010) prove that for a specific class of multiperiod control problems the affinely adjustable decision rules result in optimal adjustable solution. Bertsimas and Goyal (2010) show that the static robust solutions perform well in Stochastic Programming problems. Bertsimas et al. (2014) study cases where static decisions are worst-case optimal in two-period problems and give a tight approximation bound on the performance of static solutions, related to a measure of non-convexity of a transformation of the uncertainty set. Goyal and Lu (2014) study the performance of static solutions in problems with constraint-wise and column-wise uncertainty and provide theoretical bounds on the adaptivity gap between static and optimal adjustable solutions in such a setting.

Later, developments have been made allowing ARO to (approximately) solve problems involving adjustable integer variables. Bertsimas and Caramanis (2007) propose a sampling method for constructing adjustable robust decision rules ensuring, under certain conditions, that the robust constraints are satisfied with high probability. Bertsimas and Caramanis (2010) introduce the term of finite adaptability in two-period problems, with a fixed number of possible second-period decisions. They also show that finding the best values for these variables is NP-hard. In a later paper, Bertsimas et al. (2011a) characterize the geometric conditions for the uncertainty sets under which finite adaptability provides good approximations of the adjustable

robust solutions.

Vayanos et al. (2011) split the uncertainty set into hyper-rectangles, assigning to each of them the corresponding later-period adjustable linear and binary variables. Contrary to this, our method does not impose any geometrical form of the uncertainty subsets. Bertsimas and Georghiou (2015) propose to use piecewise linear decision rules, both for the continuous and the binary variables (for the binary variables, value 0 is implemented if the piecewise linear decision rule is positive). They use a cutting plane approach that gradually increases the fraction of the uncertainty set that the solution is robust to, reaching complete robustness when their approach terminates. In our approach, the decision rules proposed ensure full robustness after each of the so-called splitting rounds, and the more splitting rounds, the better the value of the objective function. In a recent paper, Bertsimas and Georghiou (2014) propose a different type of decision rules for binary variables. Since the resulting problems are exponential in the size of the original formulation, authors propose their conservative approximations, giving a systematic tradeoff between computational tractability and level of conservatism. In our approach, instead of imposing a functional form of the decision rules, we focus on splitting the uncertainty set into subsets with different decisions. Also, we ensure robustness precisely against the specified uncertainty set and allow non-binary integer variables.

Hanasusanto et al. (2015) apply finite adaptability to two-period decision problems with binary variables. In this setting, the decision maker can choose out of K possible decisions in the second period when the uncertain parameter value is known. For each outcome of the uncertain parameter, one of the time-2 decisions must yield a feasible solution. The optimization variables are the here-and-now decisions taken at period 1, and the set of K decisions for period 2. The resulting problems can be transformed to MILP problems of size exponential in the number K of possible decisions (in case of uncertainty in both the objective function and the constraints - for problems with uncertainty only in the objective the reformulation is polynomial). They also study the approximation quality provided by such reformulations and complexity issues. Our approach applies to general multi-period problems and allows also explicitly non-binary integer variables.

We test our methodology on problem instances from Bertsimas and Georghiou (2015) and Hanasusanto et al. (2015). The experiments reveal that our methodology performs worse on problems with uncertainty only in the objective function and on small instances, where the 'more exact' approaches of other authors can be solved fast to optimality. However, as the problems grow in size, it is able to provide comparable or better results after a significantly shorter computation.

The idea of partitioning the support of random variables in order to improve approximations of the objective function has been subject of intensive study in Stochastic

Programming (SP). There, the use of partitions is to apply some bounds on the expectation of a function of a random variable on a per-partition basis, obtaining tighter bounds in this way. Examples of such partitions are given in Birge and Wets (1986) and Frauendorfer and Kall (1988). In some cases, similarly to our methodology, these methods use dual information to decide about the positioning of the partitions (see Birge and Wets (1986)). For an overview of bounds and their partition-based refinements used in SP we refer the reader to Birge and Louveaux (1997). Despite these similarities with our approach, our method is different for its focus on the worst-case outcomes without assuming distributional information.

The composition of the remainder of the paper is as follows. Section 3.2 introduces the set-splitting methodology for the case of two-period problems with adjustable continuous variables. Section 3.3 extends the approach to multiperiod problems, and Section 3.4 extends the multiperiod case to problems with integer decision variables. Section 3.5 proposes heuristics to be used as a part of the method. Section 3.6 gives three numerical examples, showing that the methodology of our paper offers substantial gains in terms of the worst-case objective function improvement. Section 3.7 concludes and lists the potential directions for future research.

3.2 Two-period problems

For ease of exposition we first introduce our methodology on the case of two-period problems with continuous decision variables only. The extension to multi-period problems is given in Section 3.3, and the extension to problems with integer variables is given in Section 3.4.

3.2.1 Description

Consider the following two-period optimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}_1,\boldsymbol{x}_2} \quad & \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2 \\ \text{s.t.} \quad & \boldsymbol{A}_1(\boldsymbol{\zeta})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\boldsymbol{x}_2 \leq \boldsymbol{b} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{aligned} \tag{3.1}$$

where $\boldsymbol{c}_1 \in \mathbb{R}^{d_1}, \boldsymbol{c}_2 \in \mathbb{R}^{d_2}$, $\boldsymbol{b} \in \mathbb{R}^m$ are fixed parameters, $\boldsymbol{\zeta} \in \mathbb{R}^L$ is the uncertain parameter and $\mathcal{Z} \subset \mathbb{R}^L$ is a compact and convex uncertainty set. Vector $\boldsymbol{x}_1 \in \mathbb{R}^{d_1}$ is the decision implemented at time 1 before the value of $\boldsymbol{\zeta}$ is known, and $\boldsymbol{x}_2 \in \mathbb{R}^{d_2}$ is the decision vector implemented at time 2, after the value of $\boldsymbol{\zeta}$ is known. It is assumed that the functions $\boldsymbol{A}_1 : \mathbb{R}^L \to \mathbb{R}^{m\times d_1}, \boldsymbol{A}_2 : \mathbb{R}^L \to \mathbb{R}^{m\times d_2}$ are linear. We refer to the rows of matrix $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ as $\boldsymbol{a}_{1,i}^T(\boldsymbol{\zeta})$ and $\boldsymbol{a}_{2,i}^T(\boldsymbol{\zeta})$ respectively, with $\boldsymbol{a}_{1,i}(\boldsymbol{\zeta}) = \boldsymbol{P}_{1,i}\boldsymbol{\zeta}$ and $\boldsymbol{a}_{2,i}(\boldsymbol{\zeta}) = \boldsymbol{P}_{2,i}\boldsymbol{\zeta}$, where $\boldsymbol{P}_{1,i} \in \mathbb{R}^{d_1\times L}, \boldsymbol{P}_{2,i} \in \mathbb{R}^{d_2\times L}$ (uncertain parameter can

contain a single fixed component, which would result in the intercepts of the affine transformations $\boldsymbol{A}_1(\boldsymbol{\zeta}), \boldsymbol{A}_2(\boldsymbol{\zeta})$.

The static robust problem (3.1) where the decision vector $\boldsymbol{x}_2$ is independent from the value of $\boldsymbol{\zeta}$ makes no use of the fact that $\boldsymbol{x}_2$ can adjust to the revealed $\boldsymbol{\zeta}$. The adjustable version of problem (3.1) is:

$$\begin{aligned} \min_{\boldsymbol{x}_1, \boldsymbol{x}_2(\boldsymbol{\zeta}), z} \quad & z \\ \text{s.t.} \quad & \boldsymbol{c}_1^T \boldsymbol{x}_1 + \boldsymbol{c}_2^T \boldsymbol{x}_2(\boldsymbol{\zeta}) \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ & \boldsymbol{A}_1(\boldsymbol{\zeta}) \boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta}) \boldsymbol{x}_2(\boldsymbol{\zeta}) \leq \boldsymbol{b} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}. \end{aligned} \tag{3.2}$$

Since this problem is NP-hard (see Ben-Tal et al. (2009)), the concept of linear decision rules has been proposed. Then, the time 2 decision vector is defined as $\boldsymbol{x}_2 = \boldsymbol{v} + \boldsymbol{V}\boldsymbol{\zeta}$, where $\boldsymbol{v} \in \mathbb{R}^{d_2}, \boldsymbol{V} \in \mathbb{R}^{d_2 \times L}$ (see Ben-Tal et al. (2009)) and the problem is:

$$\begin{aligned} \min_{\boldsymbol{x}_1, \boldsymbol{v}, \boldsymbol{V}} \quad & z \\ \text{s.t.} \quad & \boldsymbol{c}_1^T \boldsymbol{x}_1 + \boldsymbol{c}_2^T \left(\boldsymbol{v} + \boldsymbol{V}\boldsymbol{\zeta}\right) \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ & \boldsymbol{A}_1(\boldsymbol{\zeta}) \boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta}) \left(v + V\boldsymbol{\zeta}\right) \leq \boldsymbol{b} \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}. \end{aligned} \tag{3.3}$$

In the general case such constraints are quadratic in $\boldsymbol{\zeta}$, because of the term $\boldsymbol{A}_2(\boldsymbol{\zeta}) \left(\boldsymbol{v} + \boldsymbol{V}\boldsymbol{\zeta}\right)$. Only for special cases the constraint system can be rewritten as a computationally tractable system of inequalities. Moreover, linear decision rules cannot be used if (part of) the decision vector $\boldsymbol{x}_2$ is required to be integer.

We propose a different approach. Before introducing it, we need to introduce the term of *splitting* a set. By *splitting* a set $\mathcal{Z}$ it is understood such a partition $\mathcal{Z} = \mathcal{Z}^+ \cup \mathcal{Z}^-$ that there exist $\boldsymbol{\zeta}^+ \in \mathcal{Z}^+$ and $\boldsymbol{\zeta}^- \in \mathcal{Z}^-$ such that:

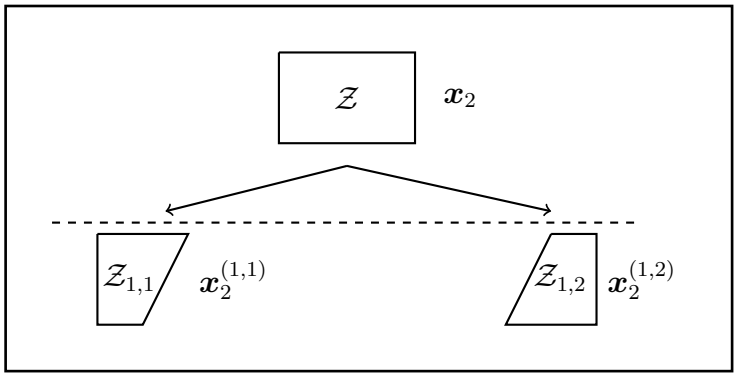
$$\boldsymbol{\zeta}^+ \in \mathcal{Z}^+ \setminus \mathcal{Z}^-, \ \boldsymbol{\zeta}^- \in \mathcal{Z}^- \setminus \mathcal{Z}^+.$$

Our idea lies in splitting the set $\mathcal{Z}$ into a collection of subsets $\mathcal{Z}_{r,s}$ where $s \in \mathcal{N}_r$ and $\cup_{s \in \mathcal{N}_r} \mathcal{Z}_{r,s} = \mathcal{Z}$ (r denotes the index of the splitting round and s denotes the set index). For each $\mathcal{Z}_{r,s}$ a different, fixed time 2 decision shall be determined. We split the set $\mathcal{Z}$ in rounds into smaller and smaller subsets using hyperplanes. In this way, all the uncertainty subsets remain convex, which is a typical assumption for RO problems. The following example illustrates this idea.

Example 3.1 *We split the uncertainty set $\mathcal{Z}$ with a hyperplane $\boldsymbol{g}^T\boldsymbol{\zeta} = h$ into the following two sets:*

$$\mathcal{Z}_{1,1} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \leq h\right\} \quad \textit{and} \quad \mathcal{Z}_{1,2} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \geq h\right\}.$$

Figure 3.1 – Scheme of the first splitting.



At time 2 the following decision is implemented:

$$\boldsymbol{x}_2 = \begin{cases} \boldsymbol{x}_2^{(1,1)} & \text{if } \boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \\ \boldsymbol{x}_2^{(1,2)} & \text{if } \boldsymbol{\zeta} \in \mathcal{Z}_{1,2} \\ \boldsymbol{x}_2^{(1,1)} \text{ or } \boldsymbol{x}_2^{(1,2)} & \text{if } \boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \cap \mathcal{Z}_{1,2}. \end{cases}$$

The splitting is illustrated in Figure 3.1. Now, the following constraints have to be satisfied:

$$\begin{cases} \boldsymbol{A}_1(\boldsymbol{\zeta})\,\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\,\boldsymbol{x}_2^{(1,1)} \leq b, & \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \\ \boldsymbol{A}_1(\boldsymbol{\zeta})\,\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\,\boldsymbol{x}_2^{(1,2)} \leq b, & \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,2}. \end{cases}$$

Since there are two values for the decision at time 2, there are also two 'objective function' values: $\boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(1,1)}$ *and* $\boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(1,2)}$*. The worst-case value is:*

$$z = \max\left\{\boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(1,1)}, \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(1,2)}\right\}.$$

After splitting $\mathcal{Z}$ *into two subsets, one is solving the following problem:*

$$\begin{array}{ll} \min & z^{(1)} \\ s.t. & \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(1,s)} \leq z^{(1)}, \quad s = 1,2 \\ & \boldsymbol{A}_1(\boldsymbol{\zeta})\,\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\,\boldsymbol{x}_2^{(1,s)} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,s}, \quad s = 1,2. \end{array} \tag{3.4}$$

Since for each s the constraint system is less restrictive than in (3.1), an improvement in the optimal value can be expected. This is illustrated in Example 3.2.

Example 3.2 *Consider the following problem in which there is only a second-stage decision vector* $\boldsymbol{x}_2 = (x_1, x_2)$ *and a single dimensional uncertain parameter* $z \in \mathcal{Z} =$

$[0,1]$*:*

$$\begin{aligned} \min\ & -x_1 - x_2 \\ s.t.\ & (1+z)x_1 \le 2 && \forall z \in \mathcal{Z} \\ & (2-z)x_2 \le 2. && \forall z \in \mathcal{Z}. \end{aligned}$$

Without adjustability we have $x_1, x_2 \le 1$ *and the optimal solution is* $x_1 = x_2 = 1$*, yielding the optimal value* -2*. However, if we split the set* $\mathcal{Z}$ *into two parts:*

$$\mathcal{Z}_1 = [0, 1/2],\ \mathcal{Z}_2 = [1/2, 1],$$

then we get:

$$\begin{aligned} \min\ & \max\{-x_1^{(1,1)} - x_2^{(1,1)}, -x_1^{(1,2)} - x_2^{(1,2)}\} \\ s.t.\ & (1+z)x_1^{(2,i)} \le 2 && \forall z_i \in \mathcal{Z}_i, i = 1,2 \\ & (2-z)x_2^{(2,i)} \le 2. && \forall z_i \in \mathcal{Z}_i, i = 1,2. \end{aligned}$$

For this optimization problem the optimal solution is given by $x_1^{(1,1)} = 4/3$*,* $x_2^{(1,1)} = 1$*,* $x_1^{(1,2)} = 1$*,* $x_2^{(1,2)} = 4/3$*, yielding the optimal value* $-7/3$*.* □

Also, the average-case performance is expected to be better than in the case of (3.1), due to the variety of time 2 decision variants.

The splitting process can be continued so that the already existing sets $\mathcal{Z}_{r,s}$ are split with hyperplanes. This is illustrated by the continuation of our example.

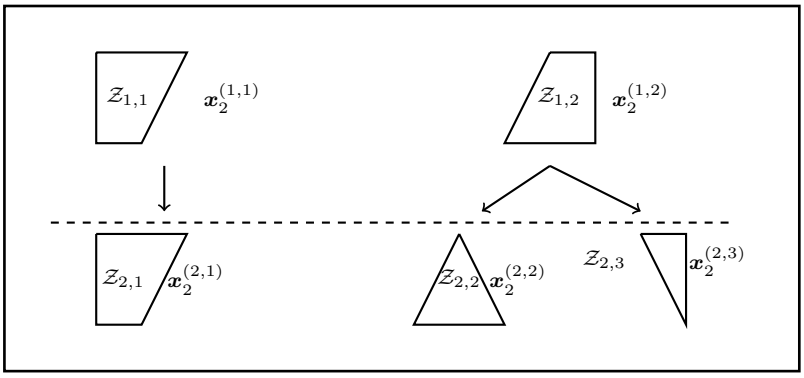
Example 3.3 *Figure 3.2 illustrates the second splitting round, where the set* $\mathcal{Z}_{1,1}$ *is not split, but the set* $\mathcal{Z}_{1,2}$ *is split with a new hyperplane into two new subsets* $\mathcal{Z}_{2,2}$ *and* $\mathcal{Z}_{2,3}$*. Then, a problem results with three uncertainty subsets and three decision variants* $\boldsymbol{x}_2^{(2,s)}$ *for time 2.* □

In general, after the r-th splitting round there are N_r uncertainty subsets $\mathcal{Z}_{r,s}$ and N_r decision variants $\boldsymbol{x}_2^{(r,s)}$. The problem is then:

$$\begin{aligned} \min\ & z^{(r)} \\ \text{s.t.}\ & \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(r,s)} \le z^{(r)}, \quad s \in \mathcal{N}_r \\ & \boldsymbol{A}_1(\boldsymbol{\zeta})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\boldsymbol{x}_2^{(r,s)} \le \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}, \qquad s \in \mathcal{N}_r = \{1, ..., N_r\}. \end{aligned} \tag{3.5}$$

The finer the splitting of the uncertainty set, the lower optimal value one may expect. In reasonable settings, as the maximum diameter of the uncertainty subsets for a

Figure 3.2 – An example of second split for the two-period case.



given r converges to 0 as $r \to +\infty$, it should hold that the optimal value of (3.5) converges to $\overline{z}_{\text{adj}}$ - the optimal value of (3.2). Indeed, Bertsimas and Caramanis (2010), who study the question of finding the optimal K variants for the time 2 decision, prove that this is true under the so-called *continuity assumption* — which, put into the terminology of this chapter, means that for any $\epsilon > 0$ and for any $\boldsymbol{z} \in \mathcal{Z}$ there exists a $\delta > 0$ and a point $(\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2)$ whose objective is ϵ-optimal in the sense of the objective function value for all $\boldsymbol{z}'$ such that $\|\boldsymbol{z} - \boldsymbol{z}'\| < \delta$. That the continuity assumption is essential for the limit to hold, it is shown in Section VI of their paper.

Determining whether further splitting is needed and finding the proper hyperplanes are crucial for an improvement in the worst-case objective value to occur. The next two subsections provide theory for determining (1) how far the current optimum is from the best possible value, (2) what are the conditions for the split to bring an improvement in the objective function value.

3.2.2 Lower bounds

As the problem becomes larger with subsequent splitting rounds, it is important to know how far the current optimal value is from $\overline{z}_{\text{adj}}$ or its lower bound. We use a lower bounding idea proposed for two-period robust problems in Hadjiyiannis et al. (2011), and used also in Bertsimas and Georghiou (2015).

Let $\overline{\mathcal{Z}} = \left\{\boldsymbol{\zeta}^{(1)}, \ldots, \boldsymbol{\zeta}^{(|\overline{\mathcal{Z}}|)}\right\} \subset \mathcal{Z}$ be a finite set of scenarios for the uncertain parameter. Consider the problem

$$\begin{aligned} \min_{w, \boldsymbol{x}_1, \boldsymbol{x}_2^{(i)}, i=1,\ldots,|\overline{\mathcal{Z}}|} \quad & w \\ \text{s.t.} \quad & \boldsymbol{c}_1^T \boldsymbol{x}_1 + \boldsymbol{c}_2^T \boldsymbol{x}_2^{(i)} \leq w, \quad i = 1, \ldots, |\overline{\mathcal{Z}}| \\ & \boldsymbol{A}_1\left(\boldsymbol{\zeta}^{(i)}\right)\boldsymbol{x}_1 + \boldsymbol{A}_2\left(\boldsymbol{\zeta}^{(i)}\right)\boldsymbol{x}_2^{(i)} \leq \boldsymbol{b}, \quad i = 1, \ldots, |\overline{\mathcal{Z}}|, \end{aligned} \tag{3.6}$$

where each $\boldsymbol{x}_1 \in \mathbb{R}^{d_1}$ and $\boldsymbol{x}_2^{(i)} \in \mathbb{R}^{d_2}$, for all i. Then, the optimal value of (3.6) is a

lower bound for $\overline{z}_{\text{adj}}$, the optimal value of (3.2) and hence, to any problem (3.5).

Since each scenario in $\overline{\mathcal{Z}}$ increases the size of the problem to solve, it is essential to include a possibly small number of scenarios determining the current optimal value of problem (3.5). The next section indicates a special class of scenarios and based on this, in Section 3.5 we propose heuristic techniques to construct $\overline{\mathcal{Z}}$.

3.2.3 How to split

In this section, we introduce the key results related to the way in which the uncertainty sets $\mathcal{Z}_{r,s}$ should be split. The main idea behind splitting the sets is as follows. For each $\mathcal{Z}_{r,s}$ we identify a finite set $\overline{\mathcal{Z}}_{r,s} \subset \mathcal{Z}_{r,s}$ of *critical scenarios*. If $\overline{\mathcal{Z}}_{r,s}$ contains more than one element, a hyperplane is constructed such that at least two elements of $\overline{\mathcal{Z}}_{r,s}$ are on different sides of the hyperplane. We call this process *dividing* the set $\overline{\mathcal{Z}}_{r,s}$. This hyperplane becomes also the splitting hyperplane of $\mathcal{Z}_{r,s}$. To avoid confusion, we use the term *split* in relation to continuous uncertainty sets $\mathcal{Z}_{r,s}$ and the term *divide* in relation to the finite sets $\overline{\mathcal{Z}}_{r,s}$ of critical scenarios.

3.2.3.1 General theorem

To obtain results supporting the decision about splitting the subsets $\mathcal{Z}_{r,s}$, we study the dual of problem (3.5). We assume that (3.5) satisfies Slater's condition. By result of Beck and Ben-Tal (2009) the dual of (3.5) is equivalent to:

$$
\begin{aligned}
\max_{\left\{\boldsymbol{\lambda}^{(r,s)}\right\}_{s=1}^{N_r},\boldsymbol{\mu}^{(r)},\left\{\left\{\boldsymbol{\zeta}^{(r,s,i)}\right\}_{s=1}^{N_r}\right\}_{i=1}^{m}} \quad & -\sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}b_i \\
\text{s.t.} \quad & \sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}\boldsymbol{a}_{1,i}\left(\boldsymbol{\zeta}^{(r,s,i)}\right)+\sum_{s\in\mathcal{N}_r}\mu_s^{(r)}\boldsymbol{c}_1=0 \\
& \sum_{i=1}^{m}\lambda_i^{(r,s)}\boldsymbol{a}_{2,i}\left(\boldsymbol{\zeta}^{(r,s,i)}\right)+\mu_s^{(r)}\boldsymbol{c}_2=0, \quad \forall s\in\mathcal{N}_r \\
& \sum_{s\in\mathcal{N}_r}\mu_s^{(r)}=1 \\
& \boldsymbol{\lambda}^{(r,s)}\geq 0, s\in\mathcal{N}_r \\
& \boldsymbol{\mu}^{(r)},\boldsymbol{\lambda}^{(r)}\geq 0 \\
& \boldsymbol{\zeta}^{(r,s,i)}\in\mathcal{Z}_{r,s}, \quad \forall s\in\mathcal{N}_r, \quad \forall 1\leq i\leq m.
\end{aligned}
\tag{3.7}
$$

Interestingly, problem (3.7) is nonconvex in the decision variables, which is not the case for duals of nonrobust problems. This phenomenom has been noted already in Beck and Ben-Tal (2009). Because Slater's condition holds, strong duality holds, and for an optimal $\overline{x}^{(r)}$ to problem (3.5), with objective value $\overline{z}^{(r)}$, there exist $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ such that the dual optimal value is attained and equal to $\overline{z}^{(r)}$. In the following, we use a shorthand notation:

$$
\boldsymbol{x}^{(r)}=\left(\boldsymbol{x}_1,\left\{\boldsymbol{x}_2^{(r,s)}\right\}_{s=1}^{N_r}\right),\boldsymbol{\lambda}^{(r)}=\left\{\boldsymbol{\lambda}^{(r,s)}\right\}_{s=1}^{N_r},\boldsymbol{\zeta}^{(r)}=\left\{\left\{\boldsymbol{\zeta}^{(r,s,i)}\right\}_{s=1}^{N_r}\right\}_{i=1}^{m},\boldsymbol{\mu}^{(r)}=(\mu_1^{(r)},\ldots,\mu_{N_r}^{(r)})^T.
$$

A similar approach is applied in the later parts of the paper. For each $s \in \mathcal{N}_r$ let us define

$$\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) = \left\{ \overline{\boldsymbol{\zeta}}^{(r,s,i)} : \quad \overline{\lambda}_i^{(r,s)} > 0 \right\},$$

which is a set of worst-case scenarios for $\boldsymbol{\zeta}$ determining that the optimal value for (3.5) cannot be better than $\overline{z}^{(r)}$. Since the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ are defined with respect to the given optimal dual solution, they are all finite.

The following theorem states that at least one of the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ for which $|\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})| > 1$ must be divided as a result of splitting $\mathcal{Z}_{r,s}$ in order for the optimal value $\overline{z}^{(r')}$ of the problem after the subsequent splitting rounds to be better than $\overline{z}^{(r)}$.

Theorem 3.1 *Assume that problem (3.5) satisfies Slater's condition, $\overline{\boldsymbol{x}}^{(r)}$ is the optimal primal solution, and $\overline{\boldsymbol{\lambda}}^{(r)}$ $\overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ is the optimal dual solution. Assume that at a splitting round $r' > r$ there exists a sequence of distinct numbers $\{j_1, j_2, ..., j_{N_r}\} \subset \mathcal{N}_{r'}$ such that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',j_s}$ for each $1 \leq s \leq N_r$, that is, each set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ remains not divided, staying a part of some uncertainty subset. Then, it holds that the optimal value $\overline{z}^{(r')}$ after the r'-th splitting round is equal to $\overline{z}^{(r)}$.* □

Proof. We construct a lower bound for the problem after the r'-th round with value $\overline{z}^{(r)}$ by choosing proper $\boldsymbol{\lambda}^{(r',s)}, \boldsymbol{\mu}^{(r')}, \boldsymbol{\zeta}^{(r',s,i)}$. Without loss of generality we assume that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',s}$ for all $s \in N_r$. We take the dual problem of the problem after the r'-th splitting round in the form (3.7). We assign the following values:

$$\lambda_i^{(r',s)} = \begin{cases} \overline{\lambda}_i^{(r,s)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_s^{(r')} = \begin{cases} \overline{\mu}_s^{(r)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases}$$

$$\boldsymbol{\zeta}^{(r',s,i)} = \begin{cases} \bar{\boldsymbol{\zeta}}^{(r,s,i)} & \text{if } s \leq N_r, \quad \overline{\lambda}_i^{(r,s)} > 0 \\ \text{any } \boldsymbol{\zeta}^{(r',s,i)} \in \mathcal{Z}_{r',s} & \text{otherwise.} \end{cases}$$

Such variables are dual feasible and give an objective value to the dual equal to $\overline{z}^{(r)}$. Since the dual objective value provides a lower bound on the primal problem after the r'-th round, the theorem follows. ■

The above result provides an important insight. If there exist sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ with more than one element each, then at least one of such sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ should be divided in the splitting process. Otherwise, by Theorem 3.1, one can construct a lower bound showing that the resulting objective value cannot improve. On the other hand, if no such $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ exists, then splitting should stop since, by Theorem 3.1, the optimal value cannot improve.

Corollary 3.1 *If for optimal* $\overline{\boldsymbol{\lambda}}^{(r,s)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ *it holds that:*

$$\left|\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})\right| \leq 1, \quad \forall s \in \mathcal{N}_r,$$

then $\overline{z}^{(r)} = \overline{z}_{adj}$*, where* $\overline{z}_{adj}$ *is the optimal value of (3.2).*

Proof. A lower-bound program with a scenario set $\overline{\mathcal{Z}} = \cup_{s\in\mathcal{N}_r}\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ has an optimal value at most $\overline{z}_{\text{adj}}$. By duality arguments similar to Theorem 3.1, the optimal value of such a lower bound problem must be equal to $\overline{z}^{(r)}$. This, combined with the fact that $\overline{z}^{(r)} \geq \overline{z}_{\text{adj}}$ gives $\overline{z}^{(r)} = \overline{z}_{\text{adj}}$. ■

Theorem 3.1 does not tell us which of the sets $\mathcal{Z}_{r,s}$ have to be split - it says only that at least one of $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ has to be split for which $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ contains more than one element. Moreover, if there exists more than one dual optimal $\overline{\boldsymbol{\lambda}}^{(r,s)}$, each of them may imply different sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ to be divided. In other words, conducting a 'proper' (in the sense of Theorem 3.1) splitting round with respect to sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$, implied by the given $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ could, in the general case, not be 'proper' with respect to sets $\overline{\mathcal{Z}}_{r,s}(\widehat{\boldsymbol{\lambda}}^{(r)})$ implied by another dual optimal $\widehat{\boldsymbol{\lambda}}^{(r)}, \hat{\boldsymbol{\zeta}}^{(r)}$. However, such a situation did not occur in any of the numerical experiments conducted in this paper.

In the following section we consider the question how to find the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ to be divided.

3.2.3.2 Finding the sets of scenarios to be divided

In this section we propose concrete methods of identifying the sets of scenarios to be divided. Such sets should be 'similar' to the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ in the sense that they should consist of scenarios $\boldsymbol{\zeta}$ that are a part of the optimal solution to the dual problem (3.7). If this condition is satisfied, such sets are expected to result in splitting decisions leading to improvements in the objective function value, in line with Theorem 3.1.

Active constraints. The first method of constructing scenario sets to be divided relies on the fact that for a given optimal solution $\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2^{(r)}$ to (3.5), a $\overline{\lambda}_i^{(r,s)} > 0$ corresponds to an active primal constraint. That means, for each $s \in \mathcal{N}_r$ we can define the set:

$$\Phi_{r,s}\left(\overline{\boldsymbol{x}}^{(r)}\right) = \left\{\boldsymbol{\zeta} : \quad \exists_i : \boldsymbol{a}_{1,i}^T(\boldsymbol{\zeta})\overline{\boldsymbol{x}}_1 + \boldsymbol{a}_{2,i}^T(\boldsymbol{\zeta})\overline{\boldsymbol{x}}_2^{(r,s)} = b_i\right\}.$$

Though some $\Phi_{r,s}\left(\overline{\boldsymbol{x}}^{(r)}\right)$ may contain infinitely many elements, one can approximate it by finding a single scenario for each constraint, solving the following problem for each s, i:

$$\begin{aligned} \min_{\boldsymbol{\zeta}} \quad & b_i - \boldsymbol{a}_{1,i}^T(\boldsymbol{\zeta})\overline{\boldsymbol{x}}_1 + \boldsymbol{a}_{2,i}^T(\boldsymbol{\zeta})\overline{\boldsymbol{x}}_2^{(r,s)} \\ & \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}. \end{aligned} \tag{3.8}$$

If for given s, i the optimal value of (3.8) is 0, we add the optimal $\boldsymbol{\zeta}$ to the set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})$. In the general case, such a set may include $\boldsymbol{\zeta}$'s for which there exists no $\overline{\lambda}_i^{(r,s)} > 0$ being a part of optimal dual solution.

Using the KKT vector of the robust problem. The active constraints approach may result in having an unnecessarily large number of critical scenarios found. Therefore, there is a need for a way to obtain the values of $\overline{\boldsymbol{\lambda}}^{(r)}$ to choose only the scenarios $\boldsymbol{\zeta}^{(r,s,i)}$ for which it holds that $\overline{\lambda}_i^{(r,s)} > 0$. This requires us to solve the problem (3.7) by solving its convex reformulation.

Here, we choose to achieve this by removing the nonconvexity of problem (3.7), which requires an additional assumption that each $\mathcal{Z}_{r,s}$ is representable by a finite set of convex constraints:

$$\mathcal{Z}_{r,s} = \{\boldsymbol{\zeta} : \quad h_{r,s,j}(\boldsymbol{\zeta}) \leq 0, \quad j = 1, ..., I_{r,s}\}, \quad \forall s \in \mathcal{N}_r, \tag{3.9}$$

where each $h_{r,s,j}(.)$ is a closed convex function. Note that this representation allows for the use of hyperplanes to split, as affine functions are also convex. For an overview of sets representable in this way we refer to Ben-Tal et al. (2015), mentioning here only that such formulation entails also conic sets. With such a set definition, by results of Gorissen et al. (2014), we can transform (3.7) to an equivalent convex problem by substituting $\lambda_i^{(r,s)}\boldsymbol{\zeta}^{(r,s,i)} = \boldsymbol{\xi}^{(r,s,i)}$. Combining this with the definition of the rows of matrices $\boldsymbol{A}_1, \boldsymbol{A}_2$, we obtain the following problem, equivalent to (3.7):

$$\begin{aligned}
\max_{\boldsymbol{\lambda}^{(r)},\boldsymbol{\mu}^{(r)},\boldsymbol{\xi}^{(r)}} \quad & -\sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}b_i \\
\text{s.t.} \quad & \sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\boldsymbol{P}_{1,i}\boldsymbol{\xi}^{(r,s,i)} + \sum_{s\in\mathcal{N}_r}\mu_s^{(r)}\boldsymbol{c}_1 = 0 \\
& \sum_{i=1}^{m}\boldsymbol{P}_{2,i}\boldsymbol{\xi}^{(r,s,i)} + \mu_s^{(r)}\boldsymbol{c}_2 = 0, \quad \forall s \in \mathcal{N}_r \\
& \sum_{s\in\mathcal{N}_r}\mu_s^{(r)} = 1 \\
& \boldsymbol{\lambda}^{(r,s)} \geq 0, s \in \mathcal{N}_r \\
& \boldsymbol{\mu}^{(r)} \geq 0 \\
& \lambda_i^{(r,s)} h_{s,j}\left(\frac{\boldsymbol{\xi}^{(r,s,i)}}{\lambda_i^{(r,s)}}\right) \leq 0, \quad \forall s \in \mathcal{N}_r, \quad i = 1, \ldots, m, \quad j = 1, ..., I_{r,s}.
\end{aligned} \tag{3.10}$$

Problem (3.10) is convex in the decision variables - it involves constraints that are either linear in the decision variables or that involve *perspective functions* of convex functions, see Boyd and Vandenberghe (2004). Optimal variables for (3.10), with substitution

$$\boldsymbol{\zeta}^{(r,s,i)} = \begin{cases} \frac{\boldsymbol{\xi}^{(r,s,i)}}{\lambda_i^{(r,s)}} & \text{for } \lambda_i^{(r,s)} > 0 \\ \boldsymbol{\zeta}^{(r,s,i)} \in \mathcal{Z}_{r,s} & \text{for } \overline{\lambda}_i^{(r,s)} = 0, \end{cases}$$

are optimal for (3.7). Hence, one may construct the sets of points to be split as:

$$\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) = \left\{ \frac{\overline{\boldsymbol{\xi}}^{(r,s,i)}}{\overline{\lambda}_i^{(r,s)}} : \quad \overline{\lambda}_i^{(r,s)} > 0 \right\}.$$

Thus, in order to obtain a set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$, one needs the solution to the convex problem (3.10). It turns out that this solution can be obtained at no extra cost apart from solving (3.5) if we assume representation (3.9) and that the tractable robust counterpart of (3.5) satisfies Slater's condition - one can use then its optimal KKT vector.

The tractable robust counterpart of (3.5), constructed using the methodology of Ben-Tal et al. (2015), is:

$$\begin{aligned} \min_{z^{(r)},\boldsymbol{x}_1,\boldsymbol{x}_2^{(r,s)},\boldsymbol{v}^{(s,i,j)},u_j^{(s,i)}} \quad & z^{(r)} \\ \text{s.t.} \quad & \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(r,s)} \leq z^{(r)}, \quad s \in \mathcal{N}_r \\ & \textstyle\sum_{j=1}^{I_{r,s}} u_j^{(s,i)} h^*_{s,i,j}\left(\frac{\boldsymbol{v}^{(s,i,j)}}{u_j^{(s,i)}}\right) \leq b_i, \quad \forall s \in \mathcal{N}_r, \forall 1 \leq i \leq m \\ & \textstyle\sum_{j=1}^{I_{r,s}} \boldsymbol{v}^{s,i,j} = \boldsymbol{P}_{1,i}^T\boldsymbol{x}_1 + \boldsymbol{P}_{2,i}^T\boldsymbol{x}_2^{(r,s)}, \quad \forall s \in \mathcal{N}_r, \forall 1 \leq i \leq m. \end{aligned} \tag{3.11}$$

Let us denote the Lagrange multipliers of the three subsequent constraint types by $\mu_s^{(r)}, \lambda_i^{(r,s)}, \boldsymbol{\xi}^{(r,s,i)}$, respectively. Now we can formulate the theorem stating that the KKT vector of the optimal solution to (3.11) gives the optimal solution to (3.10).

Theorem 3.2 *Suppose that problem (3.11) satisfies Slater's condition. Then, the components of the optimal KKT vector of (3.11) yield the optimal solution to (3.10).* □

Proof. The Lagrangian for problem (3.11) is:

$$\begin{aligned} L\left(z^{(r)}, \boldsymbol{x}^{(r)}, \boldsymbol{v}^{(s)}, \boldsymbol{u}^{(s)}, \boldsymbol{\lambda}^{(r)}, \boldsymbol{\mu}^{(r)}, \boldsymbol{\xi}^{(r)}\right) = \; & z^{(r)} + \sum_s \mu_s^{(r)} \left\{ \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(r,s)} - z^{(r)} \right\} + \\ & + \sum_{s,i} \lambda_i^{(r,s)} \left(\sum_j u_j^{(s,i)} h^*_{s,i,j}\left(\frac{\boldsymbol{v}^{s,i,j}}{u_j^{(s,i)}}\right) - b_i \right) + \\ & - \sum_{s,i} \left(\boldsymbol{\xi}^{(r,s,i)}\right)^T \left(\sum_j \boldsymbol{v}^{s,i,j} - \boldsymbol{P}_{1,i}^T\boldsymbol{x}_1 - \boldsymbol{P}_{2,i}^T\boldsymbol{x}_2^{(r,s)} \right) \end{aligned}$$

We now show that the Lagrange multipliers correspond to the decision variables with the corresponding names in problem (3.10), by deriving the Lagrange dual problem:

$$\max_{\boldsymbol{\lambda}^{(r)}\geq 0,\boldsymbol{\mu}^{(r)}\geq 0,\boldsymbol{\xi}^{(r)}} \min_{\substack{z^{(r)},\boldsymbol{x}^{(r)},\\ \boldsymbol{v}^{(s)},\boldsymbol{u}^{(s)}}} L\left(z^{(r)},\boldsymbol{x}^{(r)},\boldsymbol{v}^{(s)},\boldsymbol{u}^{(s)},\boldsymbol{\lambda}^{(r)},\boldsymbol{\mu}^{(r)},\boldsymbol{\xi}^{(r)}\right) =$$

$$= \max_{\boldsymbol{\lambda}\geq 0,\mu\geq 0,\boldsymbol{\xi}} \left\{ \min_{z^{(r)}} \left(1-\sum_s \mu_s^{(r)}\right) z^{(r)} + \min_{\boldsymbol{x}_1} \left(\sum_s \mu_s^{(r)}\boldsymbol{c}_1 + \sum_{s,i} \boldsymbol{P}_{1,i}\boldsymbol{\xi}^{(r,s,i)}\right)^T \boldsymbol{x}_1 \right.$$
$$+ \min_{\boldsymbol{x}_2^{(r,s)}} \sum_s \left(\mu_s^{(r)}\boldsymbol{c}_2 + \sum_i \boldsymbol{P}_{2,i}\boldsymbol{\xi}^{(r,s,i)}\right)^T \boldsymbol{x}_2^{(r,s)}$$
$$\left. + \sum_{s,i,j} \min_{\boldsymbol{v}^{(s,i,j)},u_j^{(s,i)}} \left\{ \lambda_i^{(r,s)} u_j^{(s,i)} h^*_{s,i,j}\left(\frac{\boldsymbol{v}^{s,i,j}}{u_j^{(s,i)}}\right) - \left(\boldsymbol{\xi}^{(r,s,i)}\right)^T \boldsymbol{v}^{s,i,j} \right\} \right\}$$

$$= \max_{\boldsymbol{\lambda}\geq 0,\boldsymbol{\mu}\geq 0,\boldsymbol{\xi}} \left\{ -\sum_s \mu_s^{(r)} b_i \,\middle|\, 1-\sum_s \mu_s^{(r)} = 0, \quad \sum_s \mu_s^{(r)}\boldsymbol{c}_1 + \sum_{s,i} \boldsymbol{P}_{1,i}\boldsymbol{\xi}^{(r,s,i)} = 0, \right.$$
$$\left. \mu_s^{(r)}\boldsymbol{c}_2 + \sum_i \boldsymbol{P}_{2,i}\boldsymbol{\xi}^{(r,s,i)} = 0, \quad \forall s, \quad \lambda_i^{(r,s)} h_{s,i,j}\left(\frac{\boldsymbol{\xi}^{(r,s,i)}}{\lambda_i^{(r,s)}}\right) \leq 0 \quad \forall s,i,j \right\}$$

Hence, one arrives at the problem equivalent to (3.10) and the theorem follows. ■

In fact, Theorem 3.2 turns out to be a special case of a the result of Beck and Ben-Tal (2009). Due to Theorem 3.2, we know that the optimal solution to (3.10), and thus to (3.7), can be obtained at no extra computational effort since most of the solvers produce the KKT vector as a part of output.

As already noted in Section 3.2.3.1, the collections of sets $\{\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})\}_{s=1}^{N_r}$ and $\{\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})\}_{s=1}^{N_r}$ may only be one of many possible collections of sets, of which at least one is to be divided. This is because different combinations of sets may correspond to different values of the optimal primal and dual variable values. Hence, there is no guarantee that even by dividing all the sets $\{\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})\}_{s=1}^{N_r}$ or $\{\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})\}_{s=1}^{N_r}$, one separates 'all the $\boldsymbol{\zeta}$ scenarios that ought to be separated'. However, the approaches presented in this section are computationally tractable and may give a good practical performance, as shown in the numerical examples of Section 3.6.

3.3 Multiperiod problems

3.3.1 Description

In this section we extend the basic two-period methodology to the case with more than two periods, which requires a more extensive notation. The uncertain parameter and the decision vector are:

$$\boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\zeta}_1 \\ \vdots \\ \boldsymbol{\zeta}_{T-1} \end{bmatrix} \in \mathbb{R}^{L_1} \times ... \times \mathbb{R}^{L_{T-1}}, \quad \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_T \end{bmatrix} \in \mathbb{R}^{d_1} \times .... \times \mathbb{R}^{d_T}.$$

Value of the component $\boldsymbol{\zeta}_t$ is revealed at time t. The decision $\boldsymbol{x}_t$ is implemented at time t, after the value of $\boldsymbol{\zeta}_{t-1}$ is known but before $\boldsymbol{\zeta}_t$ is known. We introduce a

special notation for the time-dependent parts of the vectors. The symbol $\boldsymbol{x}_{s:t}$, where $s \leq t$ shall denote the part of the vector $\boldsymbol{x}$ corresponding to periods s through t. We also define $L = \sum_{t=1}^{T-1} L_i$ and $d = \sum_{t=1}^{T} d_t$.

The considered robust multi-period problem is:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{\zeta})\boldsymbol{x} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}, \end{aligned} \tag{3.12}$$

where the matrix $\boldsymbol{A} : \mathbb{R}^L \to \mathbb{R}^{m \times d}$ is linear and its i-th row is denoted by $\boldsymbol{a}_i^T$. In the multi-period case we also split the set $\mathcal{Z}$ into a collection of sets $\mathcal{Z}_{r,s}$ where $\cup_{s \in \mathcal{N}_r} \mathcal{Z}_{r,s} = \mathcal{Z}$ for each r. By $\text{Proj}_t(\mathcal{Z}_{r,s})$ we denote the projection of the set $\mathcal{Z}_{r,s}$ onto the space corresponding to the uncertain parameters from the first t periods:

$$\text{Proj}_t(\mathcal{Z}_{r,s}) = \{\boldsymbol{\xi} : \quad \exists \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}, \quad \boldsymbol{\xi} = \boldsymbol{\zeta}_{1:t}\}.$$

Contrary to the two-period case, every subset $\mathcal{Z}_{r,s}$ shall correspond to a vector $\boldsymbol{x}^{(r,s)} \in \mathbb{R}^d$, i.e., a vector including decisions for all the periods.

In the two-period case, the time 1 decision was common for all the variants of decision variables. In the multi-period notation this condition would be written as $\boldsymbol{x}_1^{(r,s)} = \boldsymbol{x}_1^{(r,s+1)}$ for $1 \leq s \leq N_r - 1$. In the two-period case each of the uncertainty subsets $\mathcal{Z}_{r,s}$ corresponded to a separate variant $\boldsymbol{x}_2^{(r,s)}$, and given a $\boldsymbol{\zeta}$, any of them could be chosen if only it held at time 2 that $\boldsymbol{\zeta} \in \mathcal{Z}_{r,s}$. In this way, it was guaranteed that

$$\forall \boldsymbol{\zeta} \in \mathcal{Z} \quad \exists \boldsymbol{x}_2^{(r,s)} : \quad \boldsymbol{A}_1(\boldsymbol{\zeta})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\boldsymbol{x}_2^{(r,s)} \leq \boldsymbol{b}.$$

In the multi-period case the main obstacle is the need to establish *nonanticipativity constraints*, see Shapiro et al. (2009) for a discussion of nonanticipativity in the context of Stochastic Programming. Nonanticipativity requires that decisions made at time t can be based only on information available at that time.

In our context, we have that information about subsequent components of $\boldsymbol{\zeta}$ is revealed period after period, whereas at the same time decisions need to be implemented. In general up to time T one may not know to which $\mathcal{Z}_{r,s}$ the vector $\boldsymbol{\zeta}$ will *surely* belong to.

For instance, suppose that at time 1 the decision $\overline{\boldsymbol{x}}_1$ is implemented. At time 2, knowing only the value $\boldsymbol{\zeta}_1$ there may be many potential sets $\mathcal{Z}_{r,s}$ to which $\boldsymbol{\zeta}$ may belong and for which $\overline{\boldsymbol{x}}_2 = \boldsymbol{x}_2^{(r,s)}$ - all the $\mathcal{Z}_{r,s}$ for which $\boldsymbol{\zeta}_1 \in \text{Proj}_1(\mathcal{Z}_{r,s})$. Suppose that a decision $\overline{\boldsymbol{x}}_2 = \boldsymbol{x}_2^{(r,s)}$ is chosen at time 2, for some s. Then, at time 3 there must exist a set $\mathcal{Z}_{r,s}$ such that $\boldsymbol{\zeta}_{1:2} \in \text{Proj}_2(\mathcal{Z}_{r,s})$ and for which $\overline{\boldsymbol{x}}_{1:2} = \boldsymbol{x}_{1:2}^{(r,s)}$, so that its decision for time 3 can be implemented.

In general, at each time period $2 < t \leq T$ there must exist a set $\mathcal{Z}_{r,s}$ such that the vector $\boldsymbol{\zeta}_{1:t-1} \in \text{Proj}_{t-1}(\mathcal{Z}_{r,s})$, and for which it holds that $\overline{\boldsymbol{x}}_{1:t-1} = \boldsymbol{x}_{1:t-1}^{(r,s)}$, where

$\overline{\boldsymbol{x}}_{1:t-1}$ stands for the decisions already implemented. We propose an iterative splitting strategy ensuring that this postulate is satisfied.

In this strategy, the early-period decisions corresponding to various sets $\mathcal{Z}_{r,s}$ are identical, as long as it is not possible to distinguish to which of them the vector $\boldsymbol{\zeta}$ will belong. Our strategy facilitates simple determination of these equality constraints between various decisions and is based on the following notion.

Definition 1 *A hyperplane defined by a normal vector $\boldsymbol{g} \in \mathbb{R}^L$ and intercept term $h \in \mathbb{R}$ is a time t splitting hyperplane (called later t-SH) if:*

$$t = \min\left\{u : \boldsymbol{g}^T\boldsymbol{\zeta} = h \quad \Leftrightarrow \quad \boldsymbol{g}_{1:u}^T\boldsymbol{\zeta}_{1:u} = h, \quad \forall \boldsymbol{\zeta} \in \mathbb{R}^L\right\}.$$

In other words, such a hyperplane is determined by a linear inequality that depends on $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_t$, but not on $\boldsymbol{\zeta}_{t+1}, \ldots, \boldsymbol{\zeta}_T$. We shall refer to a hyperplane by the pair $(\boldsymbol{g}, h)$.

We illustrate with an example how the first splitting can be done and how the corresponding equality structure between decision vectors $\boldsymbol{x}^{(r,s)}$ is determined.

Example 3.4 *Consider a multi-period problem where $T = 3$, with a rectangular uncertainty set containing one dimensional ζ_1 and ζ_2, as depicted in Figure 3.3. We split the uncertainty set $\mathcal{Z}$ with a 1-SH $(\boldsymbol{g}, h)$. Then, two subsets result:*

$$\mathcal{Z}_{1,1} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \leq h\right\} \quad \text{and} \quad \mathcal{Z}_{1,2} = \mathcal{Z} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \geq h\right\}.$$

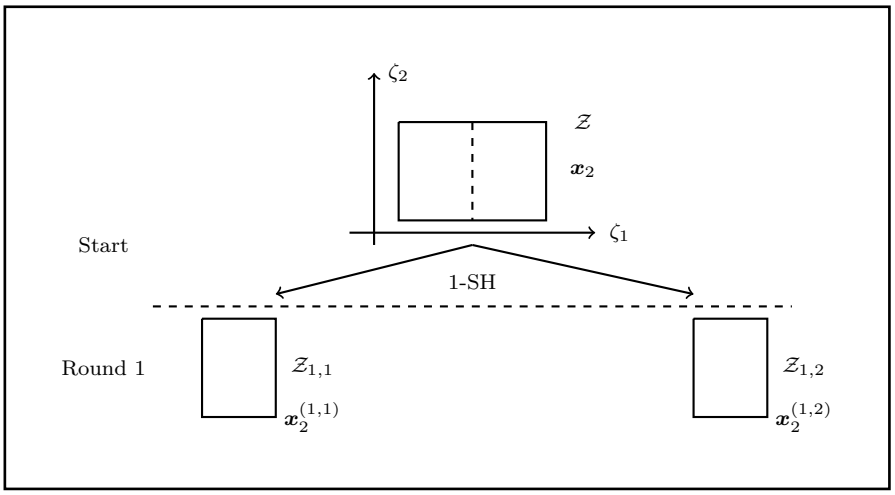
Now, there are two decision vectors $\boldsymbol{x}^{(1,1)}, \boldsymbol{x}^{(1,2)} \in \mathbb{R}^d$. Their time 1 decisions should be identical since they are implemented before the value of ζ_1 is known, allowing to determine whether $\boldsymbol{\zeta} \in \mathcal{Z}_{1,1}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{1,2}$. Thus, we add a constraint $\boldsymbol{x}_1^{(1,1)} = \boldsymbol{x}_1^{(1,2)}$. This splitting is illustrated in Figure 3.3.

The problem to be solved after the first splitting round is analogous to the two-period case, with an equality constraint added:

$$\begin{aligned}
\min_{z^{(1)}, \boldsymbol{x}^{(1,s)}} \quad & z^{(1)} \\
s.t. \quad & \boldsymbol{c}^T\boldsymbol{x}^{(1,s)} \leq z^{(1)}, \quad s = 1, 2 \\
& \boldsymbol{A}\left(\boldsymbol{\zeta}\right)\boldsymbol{x}^{(1,s)} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{1,s}, \quad s = 1, 2 \\
& \boldsymbol{x}_1^{(1,1)} = \boldsymbol{x}_1^{(1,2)}.
\end{aligned}$$

The splitting process may be continued and multiple types of t-SHs are possible. To state our methodology formally, we define a parameter $t_{max}(\mathcal{Z}_{r,s})$ for each set $\mathcal{Z}_{r,s}$. If the set $\mathcal{Z}_{r,s}$ is a result of subsequent splits with various t-SH's, the number

Figure 3.3 – A multi-period problem after a single splitting with a time-1 splitting hyperplane. Only information about ζ_1 is needed to determine if $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{1,1}$ or $\mathcal{Z}_{1,2}$ (Round 1).



$t_{max}(\mathcal{Z}_{r,s})$ denotes the largest t of them. By convention, for the set $\mathcal{Z}$ it shall hold that $t_{max}(\mathcal{Z}) = 0$. The following rule defines how the subsequent sets can be split and what the values of the parameter t_{max} for each of the resulting sets are.

Rule 3.1 *A set $\mathcal{Z}_{r,s}$ can be split only with a t-SH such that $t \geq t_{max}(\mathcal{Z}_{r,s})$. For the resulting two sets $\mathcal{Z}_{r+1,s'}, \mathcal{Z}_{r+1,s''}$ we define $t_{max}(\mathcal{Z}_{r+1,s'}) = t_{max}(\mathcal{Z}_{r+1,s''}) = t$. If the set is not split and in the next round it becomes the set $\mathcal{Z}_{r+1,s'}$ then $t_{max}(\mathcal{Z}_{r+1,s'}) = t_{max}(\mathcal{Z}_{r,s})$.* □

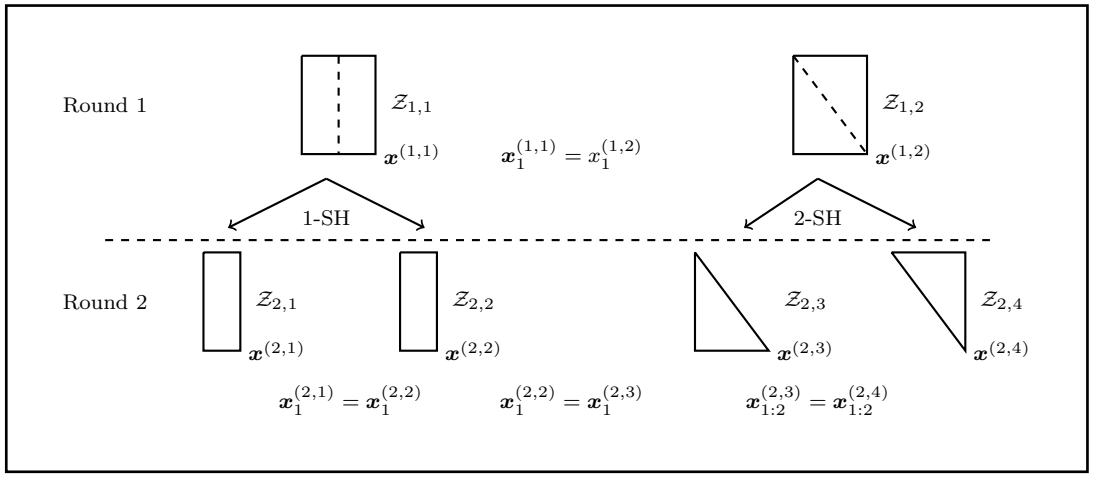
The next rule defines the equality constraints for the problem after the $(r+1)$-th splitting round, based on the problem after the r-th splitting round.

Rule 3.2 *Let a set $\mathcal{Z}_{r,s}$ be split with a t-SH into sets $\mathcal{Z}_{r+1,s'}, \mathcal{Z}_{r+1,s''}$. Then the constraint $\boldsymbol{x}_{1:t}^{(r+1,s')} = \boldsymbol{x}_{1:t}^{(r+1,s'')}$ is added to the problem after the $(r+1)$-th splitting round.*

Assume the problem after splitting round r includes sets $\mathcal{Z}^{r,s}$ and $\mathcal{Z}^{r,u}$ with a constraint $\boldsymbol{x}_{1:k_s}^{(r,s)} = \boldsymbol{x}_{1:k_s}^{(r,u)}$, and the sets $\mathcal{Z}^{r,s}, \mathcal{Z}^{r,u}$ are split into $\mathcal{Z}^{r+1,s'}, \mathcal{Z}^{r+1,s''}$ and $\mathcal{Z}^{r+1,u'}, \mathcal{Z}^{r+1,u''}$, respectively. Then, the constraint $\boldsymbol{x}_{1:k_s}^{(r+1,s'')} = \boldsymbol{x}_{1:k_s}^{(r+1,u')}$ is added to the problem after the $(r+1)$-th splitting round. □

The first part of Rule 3.2 ensures that the decision vectors $\boldsymbol{x}^{(r+1,s')}, \boldsymbol{x}^{(r+1,s'')}$ can differ only from time period $t+1$ on, since only then one can distinguish between the sets $\mathcal{Z}_{r,s'}, \mathcal{Z}_{r,s''}$. The second part of Rule 3.2 ensures that the dependence structure between decision vectors from stage r is not 'lost' after the splitting. Rule 3.2 as a whole ensures that $\boldsymbol{x}_{1:k_s}^{(r+1,s')} = \boldsymbol{x}_{1:k_s}^{(r+1,s'')} = \boldsymbol{x}_{1:k_s}^{(r+1,u')} = \boldsymbol{x}_{1:k_s}^{(r+1,u'')}$.

Figure 3.4 – Example of second splitting round for the multi-period case. Only information about ζ_1 is needed to determine whether $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{2,1}$ or $\mathcal{Z}_{2,2}$ (Round 2). However, information about both ζ_1 and ζ_2 is needed to distinguish whether $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{2,3}$ or $\mathcal{Z}_{2,4}$ (Round 2).



Rules 3.1 and 3.2 are not the only possible implementation of the splitting technique that respects the nonanticipativity restriction. However, their application in the current form does not require the decision maker to compare the sets $\mathcal{Z}_{r,s}$ for establishing the equality constraints between their corresponding decision vectors.

We illustrate the application of Rules 3.1 and 3.2 with a continuation of our example.

Example 3.5 *By Rule 3.1 we have $t_{max}(\mathcal{Z}_{1,1}) = t_{max}(\mathcal{Z}_{1,2}) = 1$. Thus, each of the sets $\mathcal{Z}_{1,1}, \mathcal{Z}_{1,2}$ can be split with a t-SH where $t \geq 1$. We split the set $\mathcal{Z}_{1,1}$ with a 1-SH and the set $\mathcal{Z}_{1,2}$ with a 2-SH. The scheme of the second splitting round is given in Figure 3.4.*

We obtain 4 uncertainty sets $\mathcal{Z}_{2,s}$ and 4 decision vectors $x^{(2,s)}$. The lower part of Figure 3.4 includes three equality constraints. The first constraint $\boldsymbol{x}_1^{(2,1)} = \boldsymbol{x}_1^{(2,2)}$ and the third constraint $\boldsymbol{x}_{1:2}^{(2,3)} = \boldsymbol{x}_{1:2}^{(2,4)}$ follow from the first part of Rule 3.2, whereas the second equality constraint $\boldsymbol{x}_1^{(2,2)} = \boldsymbol{x}_1^{(2,3)}$ is determined by the second part of Rule 3.2. The equality constraints imply that $\boldsymbol{x}_1^{(2,1)} = \boldsymbol{x}_1^{(2,2)} = \boldsymbol{x}_1^{(2,3)} = \boldsymbol{x}_1^{(2,4)}$.

The problem after the second splitting round is:

$$
\begin{aligned}
\min_{z^{(2)}, \boldsymbol{x}^{(2,s)}} \quad & z^{(2)} \\
s.t. \quad & \boldsymbol{c}^T \boldsymbol{x}^{(2,s)} \leq z^{(2)}, \quad s = 1, ..., 4 \\
& \boldsymbol{A}(\boldsymbol{\zeta})\, \boldsymbol{x}^{(2,s)} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{2,s}, \quad s = 1, ..., 4 \\
& \boldsymbol{x}_1^{(2,1)} = \boldsymbol{x}_1^{(2,2)} \\
& \boldsymbol{x}_1^{(2,2)} = \boldsymbol{x}_1^{(2,3)} \\
& \boldsymbol{x}_{1:2}^{(2,3)} = \boldsymbol{x}_{1:2}^{(2,4)}.
\end{aligned}
$$

Figure 3.5 – Time structure of the decision variants after the second splitting. Dashed horizontal lines denote the nonanticipativity (equality) constraints between decisions. The figure is motivated by Figure 3.2 in Chapter 3.1.4 of Shapiro et al. (2009).

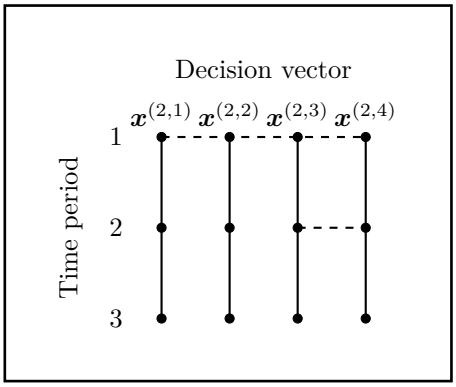
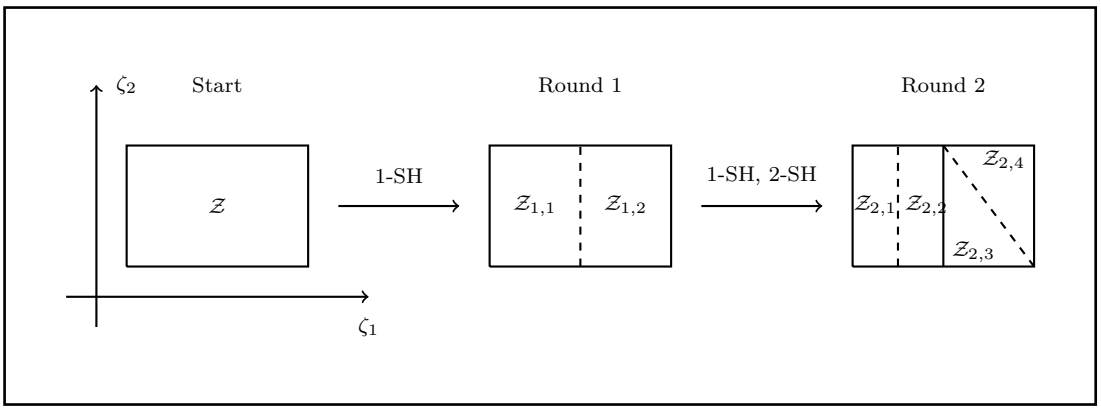


Figure 3.6 – Example of the first two splitting rounds for the multi-period case. Only information about ζ_1 is needed to determine if a point $\boldsymbol{\zeta}$ belongs to (i) $\mathcal{Z}_{1,1}$ or $\mathcal{Z}_{1,2}$ (Round 1), (ii) $\mathcal{Z}_{2,1}$ or $\mathcal{Z}_{2,2}$ (Round 2). However, information about both ζ_1 and ζ_2 is needed to distinguish whether $\boldsymbol{\zeta}$ belongs to $\mathcal{Z}_{2,3}$ or $\mathcal{Z}_{2,4}$ (Round 2).



The time structure of decisions for subsequent time periods is illustrated in Figure 3.5. Also, Figure 3.6 shows the evolution of the uncertainty set relations with the subsequent splits.

At time 1 there is only one possibility for the first decision. Then, at time 2 the value of ζ_1 is known and one can determine if $\boldsymbol{\zeta}$ is within the set $\mathcal{Z}_{1,1}$ or $\mathcal{Z}_{1,2}$, or both.

If $\boldsymbol{\zeta} \in \mathcal{Z}_{1,1}$, further verification is needed to determine whether $\boldsymbol{\zeta} \in \mathcal{Z}_{2,1}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{2,2}$, to choose the correct variant of decisions for time 2 and later.

If $\boldsymbol{\zeta} \in \mathcal{Z}_{1,2}$, the time 2 decision $\boldsymbol{x}_2^{(2,3)} = \boldsymbol{x}_2^{(2,4)}$ is implemented. Later, the value of $\boldsymbol{\zeta}_2$ is revealed and based on it, one determines whether $\boldsymbol{\zeta} \in \mathcal{Z}_{2,3}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{2,4}$. In the first case, decision $\boldsymbol{x}_3^{(2,3)}$ is implemented. In the second case, decision $\boldsymbol{x}_3^{(2,4)}$ is implemented.

If $\boldsymbol{\zeta} \in \mathcal{Z}_{1,1} \cap \mathcal{Z}_{1,2}$ (thus $\boldsymbol{\zeta}$ belongs to the tangent segment of the two sets, see Figure 3.6), then at time 2 one can implement either $\boldsymbol{x}_2^{(2,2)}$ or $\boldsymbol{x}_2^{(2,3)} = \boldsymbol{x}_2^{(2,4)}$. It is best to

choose the decision for which the entire decision vector $\boldsymbol{x}^{(r,s)}$ gives the best worst-case objective value.

If one chooses $\boldsymbol{x}_2^{(2,3)} = \boldsymbol{x}_2^{(2,4)}$, then at time 2 it is known whether $\boldsymbol{\zeta} \in \mathcal{Z}_{2,3}$ or $\boldsymbol{\zeta} \in \mathcal{Z}_{2,4}$ (or both), and the sequence of decisions for later periods is chosen. If one chooses $\boldsymbol{x}_2^{(2,2)}$ then $\boldsymbol{x}_3^{(2,2)}$ is implemented. An analogous procedure holds for other possibilities.

In general, the problem after the r-th splitting round has N_r subsets $\mathcal{Z}_{r,s}$ and decision vectors $x^{(r,s)}$. Its formulation is:

$$\begin{aligned} \min_{z^{(r)}, \boldsymbol{x}^{(r,s)}} \quad & z^{(r)} \\ & \boldsymbol{c}^T \boldsymbol{x}^{(r,s)} \leq z^{(r)}, \quad s \in \mathcal{N}_r \\ & \boldsymbol{A}\left(\boldsymbol{\zeta}\right) \boldsymbol{x}^{(r,s)} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}, \quad s \in \mathcal{N}_r \\ & \boldsymbol{x}_{1:k_s}^{(r,s)} = \boldsymbol{x}_{1:k_s}^{(r,s+1)}, \quad s \in \mathcal{N}_r \setminus \{N_r\}, \end{aligned} \tag{3.13}$$

where k_s is the number of the first time period decisions that are required to be identical for decision vectors $\boldsymbol{x}^{(r,s)}$ and $\boldsymbol{x}^{(r,s+1)}$. When Rules 3.1 and 3.2 are applied in the course of splitting, a complete set of numbers k_s is obtained from Rule 3.2 and at most $N_r - 1$ such constraints are needed. This corresponds to the sets $\{\mathcal{Z}_{r,s}\}_{s=1}^{N_r}$ being ordered in a line and having equality constraints only between the adjacent sets, see Figure 3.4, where after the second splitting round equality constraints are required only between $\boldsymbol{x}^{(2,1)}$ and $\boldsymbol{x}^{(2,2)}$, $\boldsymbol{x}^{(2,2)}$ and $\boldsymbol{x}^{(2,3)}$, and between $\boldsymbol{x}^{(2,3)}$ and $\boldsymbol{x}^{(2,4)}$.

3.3.2 Lower bounds

Similar to the two-period case, one can obtain lower bounds for the adjustable robust solution. The lower bound problem differs from the two-period case since the uncertain parameter may have a multi-period equality structure of the components that can be exploited.

Let $\overline{\mathcal{Z}} = \left\{\boldsymbol{\zeta}^{(1)}, \ldots, \boldsymbol{\zeta}^{(\left|\overline{\mathcal{Z}}\right|)}\right\} \subset \mathcal{Z}$ be a finite set of scenarios for the uncertain parameter. Then, the optimal solution to

$$\begin{aligned} \min_{w, \boldsymbol{x}_2^{(i)}, i=1,\ldots,\left|\overline{\mathcal{Z}}\right|} \quad & w \\ \text{s.t.} \quad & \boldsymbol{c}^T \boldsymbol{x}_2^{(i)} \leq w, \quad i = 1, \ldots, \left|\overline{\mathcal{Z}}\right| \\ & \boldsymbol{A}\left(\boldsymbol{\zeta}^{(i)}\right) \boldsymbol{x}^{(i)} \leq \boldsymbol{b}, \quad i = 1, \ldots, \left|\overline{\mathcal{Z}}\right| \\ & \boldsymbol{x}_{1:t}^{(i)} = \boldsymbol{x}_{1:t}^{(j)} \quad \forall_{i,j,t} : \boldsymbol{\zeta}_{1:t}^{(i)} = \boldsymbol{\zeta}_{1:t}^{(j)} \end{aligned} \tag{3.14}$$

is a lower bound for problem (3.13).

In the multi-period case it is required that for each decision vectors $\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}$ whose corresponding uncertain scenarios are identical up to time t the corresponding decisions must be the same up to time t as well (nonanticipativity restriction). This is needed since up to time t one cannot distinguish between $\boldsymbol{\zeta}^{(i)}$ and $\boldsymbol{\zeta}^{(j)}$ and the decisions made should be the same. The equality structure between the decision vectors $\boldsymbol{x}^{(i)}$ can be obtained efficiently (using at most $|\overline{\mathcal{Z}}| - 1$ vector equalities) if uncertain parameter is one-dimensional in each time period - one achieves it by sorting the set $\overline{\mathcal{Z}}$ lexicographically.

3.3.3 How to split

3.3.3.1 General theorem

We assume that (3.13) satisfies Slater's condition. By the result of Beck and Ben-Tal (2009) the dual of (3.13) is equivalent to:

$$\begin{aligned}
\max \quad & -\sum_{s\in\mathcal{N}_r}\sum_{i=1}^{m}\lambda_i^{(r,s)}b_i \\
\text{s.t.} \quad & \sum_{i=1}^{m}\lambda_i^{(r,s)}\boldsymbol{a}_i\left(\boldsymbol{\zeta}^{(r,s,i)}\right)+\mu_s^{(r)}\boldsymbol{c}+\begin{bmatrix}\nu_s^{(r)}\\0\end{bmatrix}-\begin{bmatrix}\nu_{s-1}^{(r)}\\0\end{bmatrix}=0, \quad \forall 1<s<N_r \\
& \sum_{i=1}^{m}\lambda_i^{(r,1)}\boldsymbol{a}_i\left(\boldsymbol{\zeta}^{(r,1,i)}\right)+\mu_1^{(r)}c+\begin{bmatrix}\nu_1^{(r)}\\0\end{bmatrix}=0 \\
& \sum_{i=1}^{m}\lambda_i^{(r,N_r)}\boldsymbol{a}_i\left(\boldsymbol{\zeta}^{(r,N_r,i)}\right)+\mu_{N_r}^{(r)}c-\begin{bmatrix}\nu_{r,N_r-1}^{(r)}\\0\end{bmatrix}=0 \\
& \sum_{s\in\mathcal{N}_r}\mu_s^{(r)}=1 \\
& \boldsymbol{\lambda}^{(r)},\boldsymbol{\mu}^{(r)}\geq 0 \\
& \boldsymbol{\zeta}^{(r,s,i)}\in\mathcal{Z}_{r,s}, \quad \forall s\in\mathcal{N}_r, \quad \forall 1\leq i\leq m.
\end{aligned} \tag{3.15}$$

Because of Slater's condition, strong duality holds and for an optimal primal solution $\overline{\boldsymbol{x}}^{(r)}$ with objective value $\overline{z}^{(r)}$ there exist $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\nu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ such that the optimal value of (3.15) is attained and is equal to $\overline{z}^{(r)}$. For each subset $\mathcal{Z}_{r,s}$ we define:

$$\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})=\left\{\overline{\boldsymbol{\zeta}}^{(r,s,i)}\in\mathcal{Z}_{r,s}: \quad \overline{\boldsymbol{\lambda}}_{s,i}^{(r)}>0\right\}.$$

Then, the following result holds, stating that at least one of the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$, for which $|\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})|>1$, should be split.

Theorem 3.3 *Assume that problem (3.13) satisfies Slater's condition, $\overline{\boldsymbol{x}}^{(r)}$ is the the optimal primal solution, and that $\overline{\boldsymbol{\lambda}}^{(r)}, \overline{\boldsymbol{\mu}}^{(r)}, \overline{\boldsymbol{\nu}}^{(r)}, \overline{\boldsymbol{\zeta}}^{(r)}$ are the optimal dual variables.*

Assume that at any splitting round $r' > r$ there exists a sequence of distinct numbers $\{j_1, j_2, ..., j_{N_r}\} \subset \mathcal{N}_{r'}$ such that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',j_s}$ and for each $1 \leq s \leq N_r$ it holds that $\mathcal{Z}_{r',j_s}$ results from splitting the set $\mathcal{Z}_{r,s}$. Then, the optimal value $\overline{z}^{(r')}$ is the same as $\overline{z}^{(r)}$, that is, $\overline{z}^{(r')} = \overline{z}^{(r)}$. □

Proof. We construct a lower bound for the problem after the r'-th round with value $\overline{z}^{(r)}$. Without loss of generality we assume that $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)}) \subset \mathcal{Z}_{r',s}$ for all $1 \leq s \leq N_r$. By Rules 3.1 and 3.2, the problem after the r'-th splitting round implies equality constraints $\boldsymbol{x}_{1:k_s}^{(r',s)} = \boldsymbol{x}_{1:k_s}^{(r',s+1)}$, where $1 \leq s \leq N_r - 1$. Take the dual (3.15) of the problem after the r'-th splitting round. We assign the following values for $\boldsymbol{\lambda}^{(r')}, \boldsymbol{\mu}^{(r')}$:

$$
\begin{aligned}
\lambda_i^{(r',s)} &= \begin{cases} \overline{\lambda}_i^{(r,s)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases} \\
\mu_s^{(r')} &= \begin{cases} \overline{\mu}_s^{(r)} & \text{for } 1 \leq s \leq N_r \\ 0 & \text{otherwise} \end{cases} \\
\nu_s^{(r')} &= \begin{cases} \overline{\nu}_s^{(r)} & \text{for } 1 \leq s \leq N_r - 1 \\ 0 & \text{otherwise} \end{cases} \\
\boldsymbol{\zeta}^{(r',s,i)} &= \begin{cases} \bar{\boldsymbol{\zeta}}^{(r,s,i)} & \text{if } 1 \leq s \leq N_r, \quad \overline{\lambda}_i^{(r,s)} > 0 \\ \text{any } \boldsymbol{\zeta}^{(r',s,i)} \in \mathcal{Z}_{r',s,i} & \text{otherwise.} \end{cases}
\end{aligned}
$$

These values are dual feasible and give an objective value to the dual problem equal to $\overline{z}^{(r)}$. Since the dual objective value provides a lower bound for the primal problem, the objective function value for the problem after the r'-th round cannot be better than $\overline{z}^{(r)}$. ■

Similar to the two-period case, one can prove that if each of the sets $\overline{\mathcal{Z}}_{r,s}$ has at most one element, then the splitting process may stop since the optimal objective value cannot be better than $\overline{z}^{(r)}$.

3.3.3.2 Finding the sets of scenarios to be divided

For the multi-period case, the same observations hold that have been made in the case of the two-period problem. That is, one may construct sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})$ by searching for the scenarios $\boldsymbol{\zeta}$ corresponding to active primal constraints, or sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ by using the optimal KKT variables of the tractable counterpart of (3.13). The latter approach is preferred for its inclusion only of the critical scenarios in the meaning of Theorem 3.3.

3.4 Problems with integer variables

3.4.1 Methodology

A particularly difficult application field for adjustable robust decision rules is when some of the decision variables are integer. Our methodology can be particularly useful since the decisions are fixed numbers for each of the uncertainty subset $\mathcal{Z}_{r,s}$. A general multiperiod robust adjustable problem with integer and continuous variables can be solved through splitting in the same fashion as in Section 3.2 and 3.3.

Suppose, taking the notation of Section 3.3, that the indices of components of the vector x to be integer belong to a set $\mathcal{I}$. Then, the mixed-integer version of problem (3.13) has only an additional integer condition:

$$\begin{aligned}
\min_{z^{(r)},\boldsymbol{x}^{(r,s)}} \quad & z^{(r)} \\
& \boldsymbol{c}^T\boldsymbol{x}^{(r,s)} \leq z^{(r)}, \quad s \in \mathcal{N}_r \\
& \boldsymbol{A}\left(\zeta\right)\boldsymbol{x}^{(r,s)} \leq \boldsymbol{b}, \quad \forall \zeta \in \mathcal{Z}_{r,s}, \quad s \in \mathcal{N}_r \\
& \boldsymbol{x}^{(r,s)}_{1:k_s} = \boldsymbol{x}^{(r,s+1)}_{1:k_s}, \quad i \in \mathcal{N}_r \setminus \{N_r\} \\
& \boldsymbol{x}^{(r,s)}_i \in \mathbb{Z}, \quad \forall s \in \mathcal{N}_r, \forall i \in \mathcal{I}.
\end{aligned} \tag{3.16}$$

To obtain lower bounds, we propose the analogues of the strategies given in Sections 3.2.2 and 3.3.2, with the integer condition.

3.4.2 Finding the sets of scenarios to be divided

For mixed integer optimization the available duality tools are substantially weaker than for problems with continuous variables. One can utilize the subadditive duality theorems to derive results 'similar' to the ones from Sections 3.2.3 and 3.3.3, but they are not applicable in practice. Two approaches that seem intuitively correct are: (1) separating scenarios responsible for constraints that are 'almost active' for the optimal solution $\overline{\boldsymbol{x}}^{(r)}$, (2) separating scenarios found on the basis of the LP relaxation of problem (3.16). We now discuss these two approaches.

Almost active constraints. In the continuous case, the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})$ were found by identifying $\boldsymbol{\zeta}$'s generating active constraints for the optimal primal solution. One can also apply this approach in the mixed-integer case, with a correction due to the fact that in mixed-integer problems the notion of 'active constraints' loses its proper meaning - in general case the worst-case value of a left-hand side is not a continuous function of the decision variable $\boldsymbol{x}$. For that reason, it may happen that:

$$\sup_{\boldsymbol{\zeta}\in\mathcal{Z}_{r,s}} \boldsymbol{a}_i(\boldsymbol{\zeta})^T\boldsymbol{x}^{(r,s)} < b_i,$$

even for constraints that are critical - being elements of a set of constraints prohibiting the optimal objective value of (3.16) from being better than $\overline{z}^{(r)}$. However, for each $s \in \mathcal{N}_r$ one can define an approximate set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})$ of $\boldsymbol{\zeta}$'s corresponding to 'almost active' constraints. To find such $\boldsymbol{\zeta}$'s, for a precision level $\epsilon > 0$ and $s \in \mathcal{N}_r, 1 \leq i \leq m$ one solves the following problem:

$$\begin{aligned} \min_{\zeta} \quad & b_i - \boldsymbol{a}_i \left(\boldsymbol{\zeta}\right)^T \overline{\boldsymbol{x}}^{(r,s)} - \epsilon \\ \text{s.t.} \quad & \boldsymbol{\zeta} \in \mathcal{Z}_{r,s}. \end{aligned} \tag{3.17}$$

If the result is a nonpositive optimal value, then one can add the optimal solution ζ to the set $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})$. However, this strategy may be subject to scaling problems since ϵ may imply a different degree of 'almost activeness' for different constraints. One may try to mitigate this issue by normalizing the coefficients of each constraint before solving problem (3.17).

KKT vector of the LP relaxation. Another approach for problems with integer variables, less sensitive to scaling issues, is to determine the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ corresponding to the LP relaxation of problem (3.16). This approach is expected to perform well in problems where the optimal mixed integer solution is close to the optimal solution of the LP relaxation.

Changing an integer variable by 1. Another possibility for checking the 'tightness' of a given constraint is to check whether changing the value of one (or several) of the discrete decision variables makes the constraints not hold.

The following example shows how an efficient split can be obtained by means of searching for the worst-case realizations of the uncertain parameter for the LP relaxation of the problem.

Example 3.6 *Consider the following two-stage problem, where there is only a second-stage decision vector* $\boldsymbol{x} = (x_1, x_2, x_3)$ *and a single-dimensional uncertain parameter* $z \in \mathcal{Z} = [0, 1]$*:*

$$\begin{aligned} \min\ & x_3 \\ s.t.\ & zx_1 \leq x_3 && \forall z \in \mathcal{Z} \\ & (1-z)x_2 \leq x_3 && \forall z \in \mathcal{Z} \\ & x_1 + x_2 = 1 \\ & \boldsymbol{x} \in \{0,1\}^2 \times \mathbb{R}. \end{aligned}$$

One can see this as a problem of choosing exactly one (since $x_1 + x_2 = 1$*) of the two possible ways, represented by binary choice variables* x_1*,* x_2*, from point A to point B, one of them having uncertain length* z *and the other* $1 - z$*. Without splitting,*

there are two optimal solutions $(x_1, x_2, x_3) = (0, 1, 1)$ *and* $(x_1, x_2, x_3) = (1, 0, 1)$*, with worst-case route length* 1*.*

The continuous relaxation problem has optimal solution $(x_1, x_2, x_3) = (1/2, 1/2, 1/2)$*, with optimal value 1/2. The two worst-case realizations of the uncertain parameter for the constraints in the continuous relaxation are* $z = 0$ *and* $z = 1$*. If one splits the uncertainty set separating these two points:*

$$\mathcal{Z}_1 = [0, 1/2], \ \mathcal{Z}_2 = [1/2, 1],$$

then the problem becomes:

$$\begin{array}{lll}
\min \ \max\{x_3^{(1,1)}, x_3^{(1,2)}\} & & \\
s.t. \ z x_1^{(1,i)} \leq x_3^{(1,i)} & & \forall z \in \mathcal{Z}_i, \ i = 1, 2 \\
\quad (1 - z) x_2^{(1,i)} \leq x_3^{(1,i)} & & \forall z \in \mathcal{Z}_i, \ i = 1, 2 \\
\quad x_1^{(1,i)} + x_2^{(1,i)} = 1 & & i = 1, 2 \\
\quad \boldsymbol{x}^{(1,1)}, \boldsymbol{x}^{(1,2)} \in \{0, 1\}^2 \times \mathbb{R}. & &
\end{array}$$

The optimal solution to this problem is given by $\boldsymbol{x}_2^{(1,1)} = (1, 0, 1/2)$*,* $\boldsymbol{x}_2^{(1,2)} = (0, 1, 1/2)$*, with optimal value* $1/2$*.* □

3.4.3 Problems with constraint-wise uncertainty

Some optimization problems involve constraint-wise uncertainty, that is, $\boldsymbol{\zeta}$ can be split into disjoint blocks in such a way that the data of each uncertain constraint depends on a separate block of $\boldsymbol{\zeta}$, and the uncertainty set $\mathcal{Z}$ is a direct product of uncertainty sets corresponding to the constraints (see Ben-Tal et al. (2004)). A special case are problems where uncertainty is present only in the objective function. Though in most applications this is not the case, this issue deserves a separate treatment. From Ben-Tal et al. (2004) we know that for problems with continuous decisions and constraint-wise uncertainty the optimal value obtained with adjustable decisions is equal to the one obtained with the static robust solution. However, in problems with integer decisions, adjustability may still yield an improvement in the objective function.

Up to now, we have proposed splitting the sets $\mathcal{Z}_{r,s}$ by means of dividing a set $\overline{\mathcal{Z}}_{r,s}$ containing at least two critical scenarios belonging to $\mathcal{Z}_{r,s}$. However, in case of constraint-wise uncertainty it will hold that for each constraint there is only one worst-case scenario, corresponding to a different block of $\boldsymbol{\zeta}$. Thus, splitting the uncertainty sets in order to separate the worst-case scenarios belonging to *the same* uncertainty subset cannot be applied. In such a situation, one has to resort to ad-hoc methods of finding another critical scenario within $\mathcal{Z}_{r,s}$, which may depend on the

properties of the problem at hand. We present such a heuristic approach in the route planning experiment of Section 3.6.3.

3.5 Heuristics

In this section we propose heuristics for choosing the hyperplanes to split sets $\mathcal{Z}_{r,s}$ (by splitting their corresponding sets $\overline{\mathcal{Z}}_{r,s}$) in the $(r+1)$-th splitting round, for constructing the lower bound scenario sets $\overline{\mathcal{Z}}$, and for deciding when to stop the splitting algorithm.

From now on we fix the optimal primal solution after the r-th splitting round $\overline{\boldsymbol{x}}^{(r)}$ and the sets $\overline{\mathcal{Z}}_{r,s}$, making no distinction between the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r)})$ obtained by using the optimal KKT vector of the problems' LP relaxations and the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{\lambda}}^{(r)})$ obtained by searching constraint-wise for scenarios that make the constraints (almost) active. We only consider splitting of sets $\mathcal{Z}_{r,s}$ for which $|\overline{\mathcal{Z}}_{r,s}| > 1$.

3.5.1 Choosing the t for the t-SHs

In multi-period problems one must determine the t for the t-SH, and this choice should balance two factors. Intuitively, the set $\mathcal{Z}_{r,s}$ should be split with a $t \geq t_{max}(\mathcal{Z}_{r,s})$ for which the components $\boldsymbol{\zeta}_t$ are most dispersed over $\boldsymbol{\zeta} \in \overline{\mathcal{Z}}_{r,s}$. On the other hand, choosing a high value of t in an early splitting round reduces the range of possible t-SHs in later rounds because of Rule 3.1.

We propose that each $\mathcal{Z}_{r,s}$ is split with a t-SH for which the components $\boldsymbol{\zeta}_t$ show biggest dispersion within the set $\overline{\mathcal{Z}}_{r,s}$ (measured, for example, with variance) and where $t_{max}(\mathcal{Z}_{r,s}) \leq t \leq t_{max}(\mathcal{Z}_{r,s}) + q$, with q being a predetermined number. If the dispersion equals 0 for all $t_{max}(\mathcal{Z}_{r,s}) \leq t \leq t_{max}(\mathcal{Z}_{r,s}) + q$ then we propose to choose the smallest $t \geq t_{max}(\mathcal{Z}_{r,s})$ such that the components $\boldsymbol{\zeta}_t$ show a nonzero dispersion within $\overline{\mathcal{Z}}_{r,s}$.

3.5.2 Splitting hyperplane heuristics

In this subsection we provide propositions for constructing the splitting hyperplanes.

Heuristic 3.1 *The idea of this heuristic is to determine the two most distant scenarios in $\overline{\mathcal{Z}}_{r,s}$ and to choose a hyperplane that separates them strongly.*

Find the $\boldsymbol{\zeta}^{(a)}, \boldsymbol{\zeta}^{(b)} \in \overline{\mathcal{Z}}_{r,s}$ maximizing $\left\| \boldsymbol{\zeta}^{(i)}_{1:t} - \boldsymbol{\zeta}^{(j)}_{1:t} \right\|_2$ over $\boldsymbol{\zeta}^{(i)}, \boldsymbol{\zeta}^{(j)} \in \overline{\mathcal{Z}}_{r,s}$. Then, split the set $\mathcal{Z}_{r,s}$ with a t-SH defined by:

$$\boldsymbol{g}_j = \begin{cases} \boldsymbol{\zeta}^{(a)}_j - \boldsymbol{\zeta}^{(b)}_j & \text{if } j \leq t \\ 0 & \text{otherwise} \end{cases}, \quad h = \frac{\boldsymbol{g}^T \left(\boldsymbol{\zeta}^{(a)} + \boldsymbol{\zeta}^{(b)} \right)}{2}.$$

If (3.8) or (3.17) is used to find critical binding scenarios, then these problems could have multiple binding scenarios. Then, the separation of optimal facets may yield better results than of a single $\boldsymbol{\zeta}$ found to be optimal for (3.8), (3.17). Then, the heuristic would separate the two most distant facets with, for example, their bisector hyperplane.

Heuristic 3.2 *The idea of this heuristic is to divide the set $\overline{\mathcal{Z}}_{r,s}$ into two sets whose cardinalities differ by as little as possible.*

Choose an arbitrary normal vector g for the t-SH. Then, determine the intercept term h such that the term $||\overline{\mathcal{Z}}^{-}_{r,s}| - |\overline{\mathcal{Z}}^{+}_{r,s}||$ is minimized, with

$$\overline{\mathcal{Z}}^{-}_{r,s} = \overline{\mathcal{Z}}_{r,s} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \leq h\right\}, \quad \overline{\mathcal{Z}}^{+}_{r,s} = \overline{\mathcal{Z}}_{r,s} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \geq h\right\}.$$

The best h can be found using binary search.

Heuristic 3.3 *The idea of this heuristic is to split the set $\mathcal{Z}_{r,s}$ with a hyperplane, and to manipulate the late period decisions while keeping the early-period decisions fixed, in such a way that the maximum worst-case 'objective function' for the two sets is minimized. We describe it for the multi-period case.*

Choose an arbitrary normal vector g for the t-SH. For a given intercept h define the two sets:

$$\mathcal{Z}^{h-}_{r+1,s} = \mathcal{Z}_{r,s} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \leq h\right\}, \quad \mathcal{Z}^{h+}_{r+1,s} = \mathcal{Z}_{r,s} \cap \left\{\boldsymbol{\zeta} : \boldsymbol{g}^T\boldsymbol{\zeta} \geq h\right\}.$$

For a fixed $\boldsymbol{g}$ we define the following function (note that the formulation only includes the constraints related to the given s):

$$\begin{aligned}
\tau(h) = \min_{\boldsymbol{x}^{(r,s')}, \boldsymbol{x}^{(r,s'')}, w} \quad & w \\
s.t. \quad & \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(r,s')} \leq w \\
& \boldsymbol{c}_1^T\boldsymbol{x}_1 + \boldsymbol{c}_2^T\boldsymbol{x}_2^{(r,s'')} \leq w \\
& \boldsymbol{A}_1(\boldsymbol{\zeta})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\boldsymbol{x}_2^{(r,s')} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}^{h-}_{r+1,s} \\
& \boldsymbol{A}_1(\boldsymbol{\zeta})\boldsymbol{x}_1 + \boldsymbol{A}_2(\boldsymbol{\zeta})\boldsymbol{x}_2^{(r,s'')} \leq \boldsymbol{b}, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z}^{h+}_{r+1,s} \\
& \boldsymbol{x}^{(r,s')}_{1:t_{max}(\mathcal{Z}_{r,s})} = \boldsymbol{x}^{(r,s'')}_{1:t_{max}(\mathcal{Z}_{r,s})} = \overline{\boldsymbol{x}}^{(r,s)}_{1:t_{max}(\mathcal{Z}_{r,s})}.
\end{aligned} \tag{3.18}$$

Equality constraints ensure that the decision variables related by equality constraints to other decision vectors stay with the same values (not to lose the feasibility of the decision vectors for sets $\mathcal{Z}_{r,p}$, where $p \neq s$). The aim is to minimize $\tau(h)$ over the domain of h for which both $\mathcal{Z}^{h-}_{r+1,s}$ and $\mathcal{Z}^{h+}_{r+1,s}$ are nonempty. Function $\tau(h)$ is quasiconvex in h, which has been noted in a different setting in Bertsimas et al. (2010).

3.5.3 Constructing the lower bound scenario sets

The key premise is that the size of the set $\overline{\mathcal{Z}}^{(r)}$ (the lower bound scenario set after the r-th splitting round) should be kept limited since each additional scenario increases the size of the lower bound problem. Hence, it is important that the limited number of scenarios covers set $\mathcal{Z}$ well.

Summing the scenario sets. One approach is to use $\overline{\mathcal{Z}}^{(r)} = \cup_{s \in \mathcal{N}_r} \overline{\mathcal{Z}}_{r,s}$ after each splitting round, since $\overline{\mathcal{Z}}_{r,s}$ approximates the set of the scenarios that are part of the current dual optimal solution, yielding a bound on the optimal value of the objective function.

To reduce the size of $\overline{\mathcal{Z}}^{(r)}$, we propose that $\overline{\mathcal{Z}}$ contains at most k elements of each $\overline{\mathcal{Z}}_{r,s}$, where k is a predetermined number. This approach implies that the lower bound sequence $\left\{\overline{w}^{(r)}\right\}$, where $\overline{w}^{(r)}$ is the optimal value of the lower bound problem after the r-th splitting round, needs not be nondecreasing in r.

Incremental building of a scenario set. To ensure a nondecreasing lower bound sequence, one can construct the sets incrementally, starting with $\overline{\mathcal{Z}}^{(1)}$ after the first splitting round and enlarging it with new scenarios after each splitting round. We describe a possible variant of this idea for the multi-period case.

Assume that problem (3.14) has been solved after the r-th splitting round, the lower-bounding scenario set is $\overline{\mathcal{Z}}^{(r)}$, the optimal value of the lower-bounding problem is $\overline{w}^{(r)}$, and $\overline{\boldsymbol{x}}^{(i)}$, $i = 1, \ldots, |\overline{\mathcal{Z}}^{(r)}|$, are the decision vectors from the lower bound problem after the r-th splitting round. Suppose that after the $(r+1)$-th splitting round there is a candidate scenario $\boldsymbol{\zeta}' \in \overline{\mathcal{Z}}_{r+1,s}$ for being added to the lower-bound scenario set $\overline{\mathcal{Z}}^{(r+1)}$. Then, scenario $\boldsymbol{\zeta}'$ is added if (1) there is no $1 \leq i \leq |\overline{\mathcal{Z}}^{(r)}|$ such that $A(\boldsymbol{\zeta}')\left(\overline{\boldsymbol{x}}^{(i)}\right) \leq b$, (2) there exists no $\boldsymbol{x}^{(\zeta')}$ such that the optimal value to the problem:

$$
\begin{aligned}
\max_{\kappa, \boldsymbol{x}^{(\zeta')}} \quad & \kappa \\
\text{s.t.} \quad & \boldsymbol{c}^T \boldsymbol{x}^{(\zeta')} \leq \overline{w}^{(r)} - \kappa \\
& A(\boldsymbol{\zeta}')\overline{\boldsymbol{x}}^{(i)} \leq b, \quad \forall i \\
& \boldsymbol{x}^{(\zeta')}_{1:t} = \boldsymbol{x}^{(i)}_{1:t} \quad \forall 1 \leq i \leq |\overline{\mathcal{Z}}^{(r)}|, \quad \forall t : \boldsymbol{\zeta}'_{1:t} = \boldsymbol{\zeta}^{(i)}_{1:t},
\end{aligned}
$$

is nonnegative. Condition (1) excludes the case when there exists already $\boldsymbol{\zeta}^{(i)} \in \overline{\mathcal{Z}}^{(r)}$ whose corresponding decision vector $\boldsymbol{x}^{(i)}$ is robust to $\boldsymbol{\zeta}'$. Condition (2) excludes the case when it is possible to construct a decision vector for $\boldsymbol{\zeta}'$ satisfying the nonanticipativity constraints in relation to decision vectors corresponding to $\boldsymbol{\zeta} \in \overline{\mathcal{Z}}^{(r)}$, and yielding an objective value $\boldsymbol{c}^T \boldsymbol{x}^{(\zeta')} \leq \overline{w}^{(r)}$. Such a scenario brings no value as it is known that a lower bound obtained using $\boldsymbol{\zeta}'$ in addition to $\overline{\mathcal{Z}}^{(r)}$ would be at most equal to the lower bound obtained using only $\overline{\mathcal{Z}}^{(r)}$.

Simple heuristic. We propose also an approach that combines approximately the properties of the two propositions above and is fast at the same time. The idea is to build up the lower-bounding set iteratively and add from each $\overline{\mathcal{Z}}_{r,s}$ the k scenarios whose sum of distances from the elements of $\overline{\mathcal{Z}}^{(r-1)}$ is largest. The distance between two vectors is measured by the 2-norm.

3.5.4 Stopping the algorithm

As the splitting continues, the computational workload related to solving the split problem grows because of the number of variables and uncertainty subsets. We propose three stopping rules for the splitting method: (1) when the objective value is closer to the lower bound than a predetermined threshold level, (2) when the limit of total computational time is reached, (3) when the maximum number of splitting rounds is reached.

3.6 Numerical experiments

3.6.1 Capital budgeting

The first numerical experiment involves no fixed recourse and is the capital budgeting problem taken from Hanasusanto et al. (2015). In this problem, a company can allocate an investment budget of B to a subset of projects $i \in \{1, \ldots, N\}$. Each project i has uncertain costs $c_i(\zeta)$ and uncertain profits $r_i(\boldsymbol{\zeta})$, modelled as affine functions of an uncertain vector $\boldsymbol{\zeta}$ of risk factors. The company can invest in a project before or after observing the risk factors $\boldsymbol{\zeta}$. A postponed investment in project i incurs the same costs $c_i(\boldsymbol{\zeta})$, but yields only a fraction $\theta \in [0, 1)$ of the profits $r_i(\boldsymbol{\zeta})$.

The problem of maximizing the worst-case return can be formulated as:

$$
\begin{aligned}
\max \quad & R \\
\text{s.t.} \quad & R \leq \boldsymbol{r}(\boldsymbol{\zeta})^T(\boldsymbol{x} + \theta \boldsymbol{y}), \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\
& \boldsymbol{c}(\boldsymbol{\zeta})^T(\boldsymbol{x} + \boldsymbol{y}) \leq B, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\
& \boldsymbol{x} + \boldsymbol{y} \leq 1 \\
& \boldsymbol{x}, \boldsymbol{y} \in \{0, 1\}^N,
\end{aligned}
$$

where the decisions x_i and y_i attain value 1 if and only if an early or late investment in project i is undertaken, respectively. The uncertainty set is $\mathcal{Z} = [-1, 1]^F$, where F is the number of risk factors.

We adopt the same random data setting as Hanasusanto et al. (2015). In all instances

Table 3.1 – Results of Hanasusanto et al. (2015) for the capital budgeting problem. K is the number of time-2 decision variants allowed and N is the number of projects. The columns are (1) - percentage of instances solved to optimality within 2h, (2) - average solution time of the instances solved within 2h, (3) - average objective improvements (including the suboptimal solutions from Gurobi for the instances not solved within 2h.

	$K=2$			$K=3$			$K=4$		
N	(1) (%)	(2) s	(3) (%)	(1) (%)	(2) s	(3) (%)	(1) (%)	(2) s	(3) (%)
5	100	¡1	48.67	100	1	68.71	100	36	79.50
10	100	4	59.34	74	1210	86.91	0	-	102.48
15	100	512	63.69	0	-	91.75	0	-	106.93
20	2	5232	64.78	0	-	93.20	0	-	108.61
25	0	-	64.85	0	-	93.72	0	-	109.10
30	0	-	64.98	0	-	94.08	0	-	109.42

we use $F = 4$. The project costs and profits are modelled as:

$$c_i(\boldsymbol{\zeta}) = (1 + \boldsymbol{\Phi}_i^T \boldsymbol{\zeta}/2)c_i^0, \qquad r_i(\boldsymbol{\zeta}) = (1 + \boldsymbol{\Psi}_i^T \boldsymbol{\zeta}/2)r_i^0, \quad i = 1, \ldots, N.$$

Parameters c_i^0 and r_i^0 are the nominal costs and profits of project i, whereas Φ_i and Ψ_i represent the i-th rows of the factor loading matrices $\boldsymbol{\Phi}, \boldsymbol{\Psi} \in \mathbb{R}^{N\times 4}$ as column vectors. The nominal costs c^0 are sampled uniformly from $[0, 10]^N$, and the nominal profits are set to $r^0 = c^0/5$. The components in each row of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are sampled uniformly from the unit simplex in $\mathbb{R}^4$. The investment budget is set to $B = 1^T c^0/2$, and we set $\theta = 0.8$. Table 3.1 gives the results of Hanasusanto et al. (2015), who apply a K-adaptability approach and sample 100 instances for each combination of N and K (the number of time-2 decision variants) and try to solve it to optimality within a time limit of 2h per instance.

We sample 50 instances for each N and conduct 8 splitting rounds for $N = 5, 10$, 6 for $N = 15, 20$ and 4 for $N = 25, 30$ (for smaller problems one can allow more splitting rounds to obtain better objectives and still operate within reasonable time limits). To split the uncertainty sets we use the worst-case scenarios coming from the optimal KKT vector of the LP relaxation of the robust MILP problems (see Section 3.2.3.2). In each splitting round we split all subsets $\mathcal{Z}_{r,s}$ for which $|\overline{\mathcal{Z}}_{r,s}| > 1$. The splitting hyperplanes are constructed using Heuristic 3.1 (see Section 3.5.2). The upper bound scenario sets are constructed according to the 'simple heuristic' (see Section 3.5.3) with $k = 2$. The after-splitting robust MILP problems are solved with Gurobi precision set to 0.5%. All problems were formulated using CVX package and solved with Gurobi solver on an Intel Core 2.66GHz computer.

Apart from the worst-case results, for each instance we conduct a simulation study by sampling from $[-1, 1]^4$ uniformly 500 scenarios of the risk factors' values and computing the objective function values obtained using the static robust solution and our splitting-based adjustable solution.

Table 3.2 – Our results for the capital budgeting problem. 'Splitting rounds' denotes the number of splitting rounds conducted. 'Initial gap' is the optimality gap for the static robust solution and the lower bound obtained after the first splitting round. 'Final gap' is the optimality gap computed with the objective value and lower bound after the last splitting round. 'Average case improvement' denotes the increase of the average-case objective value obtained with the adjustable decisions, relative to the one yielded by the static solution. The relative optimality gaps are computed as $\frac{(\text{UB}-\text{LB})}{0.5(\text{UB}+\text{LB})} * 100\%$, where LB is the objective function value and UB is the upper bound value.

Splitting rounds	N	Obj improvement (%)	Initial gap (%)	Final gap (%)	Average case improvement (%)	Mean time (s)
8	5	57.89	106.09	39.00	12.11	5.40
	10	93.81	100.15	27.68	20.36	26.81
6	15	102.63	100.00	24.29	23.13	4.72
	20	107.81	100.00	22.24	24.79	5.43
4	25	105.33	100.00	23.48	24.30	3.96
	30	106.88	100.00	22.93	24.80	6.54

Table 3.2 gives the results of our methodology. All the instances have been solved fast, with the largest average time equal to 26.81s. We remark here that, typically for problems with binary variables, the distribution of the solution times is heavy-tailed, and whereas most of the instances are solved within 2-3s, some instances take much more time and influence the average times in this way. Our methodology performs worse on the small instances, which the 'more exact' method of Hanasusanto et al. (2015) can solve efficiently in short time. For larger instances our improvements in the objective value are close to the best values of Hanasusanto et al. (2015) for larger instances $N = 20, 25, 30$ - ours being $107.81, 105.33, 106.88\%$ versus their $108.61, 109.10, 109.42\%$, respectively.

We also compare the running time performance of our method to the results of Hanasusanto et al. (2015) though we should mention that the main objective of Hanasusanto et al. (2015) was to find the best solution using a fixed number of time 2 policies. For larger instances ($N \geq 15$) the results of Hanasusanto et al. (2015) are based on suboptimal solutions from Gurobi obtained after 2 hours of computation per instance (see Table 3.1), whereas our method uses on average less than 27s per instance, with most of the mean times being less than 7s. Upon request, we obtained the Gurobi output of Hanasusanto et al. (2014). It reveals that in majority of instances studied by them, the objective value obtained by the solver after 60s is within 5% of the end objective value obtained after the time limit of 7200s, given in Table 3.1.

The right part of Table 3.2 gives the average-case improvements obtained using the adjustable decisions. The improvements are significantly smaller than the worst-case improvements, stabilizing around the level of 25% for larger N.

Figure 3.7 – Capital budgeting problem. Plots of initial and final upper bound on the worst-case objective function values and the initial and final worst-case objective function values (average over all problem instances for each N).

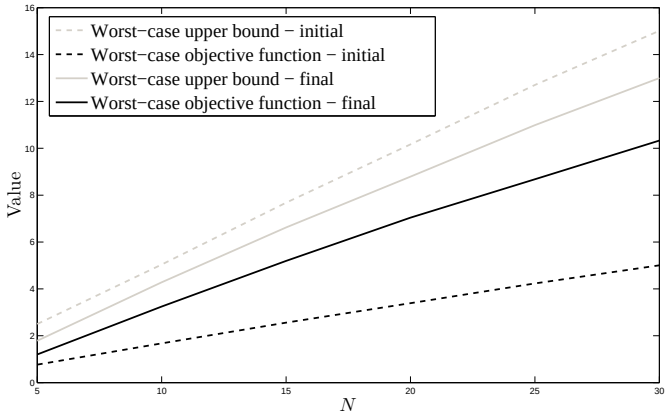


Figure 3.7 shows the average (over problem instances for given N) improvements of the worst-case objective functions and the upper bounds for all N. One can see that the relative gap between the upper bound values and the worst-case objective values decreases significantly with the number N of projects.

We summarize now the results of the first numerical example. Hanasusanto et al. (2015) give good worst-case objective value improvements with a small number of time-2 decision variants (at most 4) after a longer computation time, whereas our splitting method gives such improvements after a short computation time, but with more time-2 decision variants. For example, 9 splitting rounds typically result in a division of the uncertainty set $\mathcal{Z}$ into more than 10 parts, each with a corresponding time-2 decision variant. Thus, our methodology is preferred when it is the computation time, and not the number of decision variants, that is to be kept low.

3.6.2 Lot sizing

As the second numerical experiment we consider a multi-stage lot sizing problem taken from Bertsimas and Georghiou (2015). The problem entails a single product, T time periods, and the following parameters:

- ζ_t, where $t = 1, \ldots, T$, is the uncertain demand in period t
- l_t, where $t = 2, \ldots, T$, is the lowest possible demand in period t
- u_t, where $t = 2, \ldots, T$, is the highest possible demand in period t
- c_{y_n}, where $n = 1, \ldots, N$, is the per product unit of buying a fixed quantity q_n

- c_x is the ordering cost per product unit for purchases that are delivered in the subsequent period
- c_h is the holding cost per product unit
- $\overline{x}_{\text{tot},t}$, where $t = 2, \ldots, T$, is the cumulative orders limit up to time period t.

The variables are:

- I_t, where $t = 1, \ldots, T$, is the level of available inventory after period t
- x_t, where $t = 1, \ldots, T-1$ is the product amount ordered in period t, after $\zeta_1, \ldots, \zeta_t$ is known, and delivered in period $t+1$, at unit price c_x
- y_{nt}, where $n = 1, \ldots, N$, $t = 2, \ldots, T$, is a binary decision made after $\zeta_1, \ldots, \zeta_t$ is known, whether to buy a fixed quantity q_n of the product in time period t, delivered in the same time period.

The difference between the ordering decisions x_t and y_{nt} is thus that x_t stands for continuous ordering decisions that result in products being delivered with a delay of one time period, and y_{nt} stands for a fixed-size product amount delivered immediately.

The problem is to minimize the worst-case combined ordering and holding costs (referred later to as the 'total cost'), subject to cumulative ordering constraints:

$$
\begin{aligned}
\min \quad & z \\
\text{s.t.} \quad & \sum_{t=2}^{T} \left(c_x x_{t-1}(\boldsymbol{\zeta}_{1:t-1}) + c_h I_t(\boldsymbol{\zeta}_{1:t}) + \textstyle\sum_{n=1}^{N} c_{y_n} q_n y_{nt}(\boldsymbol{\zeta}_{1:t}) \right) \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\
& \left.
\begin{array}{l}
I_t(\boldsymbol{\zeta}_{1:t}) = I_{t-1}(\boldsymbol{\zeta}_{1:t-1}) + x_{t-1}(\boldsymbol{\zeta}_{1:t-1}) + \sum_{n=1}^{N} q_n y_{nt}(\boldsymbol{\zeta}_{1:t}) - \zeta_t \\
0 \leq x_{t-1}(\boldsymbol{\zeta}_{1:t-1}) \\
0 \leq I_t(\boldsymbol{\zeta}_{1:t}) \\
\sum_{j=1}^{t-1} x_j(\boldsymbol{\zeta}_{1:j}) \leq \overline{x}_{\text{tot},t}
\end{array}
\right\} \quad \begin{array}{c} \forall t = 2, \ldots, T, \\ \forall \boldsymbol{\zeta} \in \mathcal{Z} \end{array} \\
& y_{nt}(\boldsymbol{\zeta}_{1:t}) \in \{0, 1\}, \quad \forall n, t \\
& x_t(\boldsymbol{\zeta}_{1:t}) \geq 0, \quad \forall t,
\end{aligned}
\tag{3.19}
$$

where

$$
\mathcal{Z} = \{\boldsymbol{\zeta} : \zeta_1 = 1, \quad l_t \leq \zeta_t \leq u_t, \quad t = 2, \ldots, T\}.
$$

The above problem is transformed by eliminating the variables I_t for $t = 2, \ldots, T$. The adjustable variables are x_t, allowed to depend on $\boldsymbol{\zeta}_{1:t}$ for $t = 1, \ldots, T-1$ and y_{nt}, allowed to depend on $\boldsymbol{\zeta}_{1:t}$ for $t = 2, \ldots, T$.

Table 3.3 – Results of Bertsimas and Georghiou (2015) for the lot sizing problem. The relative optimality gaps are computed as $\frac{(\text{UB}-\text{LB})}{0.5(\text{UB}+\text{LB})} * 100\%$, where UB is the objective function value and LB is the lower bound value. 'Nonadaptive gap' denotes the relative optimality gap computed for the solution where the integer decisions are static and the linear decision rules are implemented for the continuous decision variables. '$\mathcal{PB}_t(1)$ gap' denotes relative optimality gap computed for the solution obtained using the binary adjustability technique used by the authors and where the linear decision rules are implemented for the continuous decision variables. 1% and 5% at the top of the Table are two variants of solver precision used when solving the MILP problems.

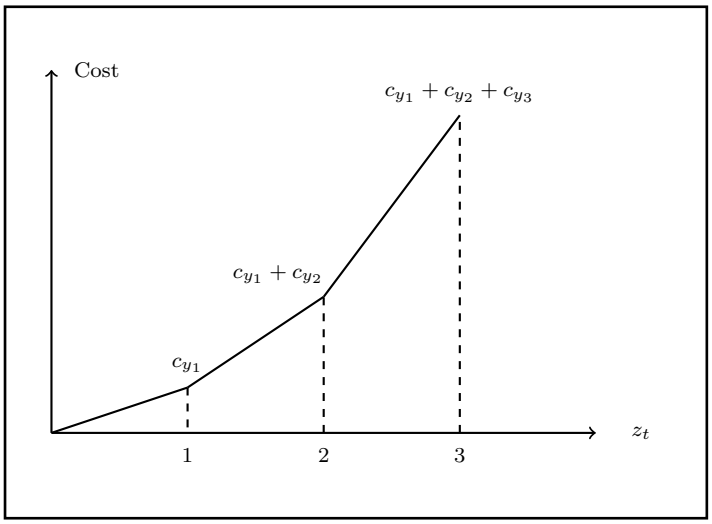
		1% optimality			5% optimality		
N	T	$\mathcal{PB}_t(1)$ gap (%)	Nonadaptive gap (%)	Mean time (s)	$\mathcal{PB}_t(1)$ gap (%)	Nonadaptive gap (%)	Mean time (s)
2	2	0	17.6	0.1	0.6	17.6	0.4
	4	24.2	68.6	50.6	27.3	68.6	45.5
	6	37.4	62.0	4833.8	38.9	62.1	956.8
	8	37.9	84.4	27531.1	38.0	84.4	19573.1
	10	39.7	89.9	35716.6	42.0	89.9	31464.1
3	2	0	27.6	0.1	1.2	27.6	0.1
	4	17.2	73.3	3381.8	23.9	73.3	781.6
	6	34.5	66.2	9181.0	38.4	66.1	3298.1
	8	37.6	83.4	28742.7	38.1	83.7	21885.5
	10	-	89.7	-	41.1	90.7	39141.5

Problem parameters are sampled as in Bertsimas and Georghiou (2015). Ordering costs are chosen from $c_x \in [0; 5]$ and $c_{y_n} \in [0; 10]$, separately for all $n = 1, \ldots, N$, such that $c_x < c_{y_n}$. In this way, the per-item costs of the fixed-size lots of products is always higher than of the product amounts ordered in continuous decisions, and hence, the only advantages of fixed-size lots are in their immediate delivery.

Holding costs are elements of $c_h \in [0; 10]$ with the fixed ordering quantities set to $q_n = 100/N$ for all $n = 1, \ldots, N$. The cumulative ordering budget is set to $\overline{x}_{\text{tot},t} = \sum_{s=1}^{t-1} \bar{x}_s$ for $t = 2, \ldots, T$, with $\bar{x}_t \in [0; 100]$ and the lower and upper bounds for the demand are sampled uniformly as $l_t \in [0; 25]$ and $u_t \in [75; 100]$, $t = 2, \ldots, T$. We assume that the initial inventory level I_1 equals zero. Table 3.3 gives the results obtained by Bertsimas and Georghiou (2015) using their methodology of piecewise linear decision rules for the decision variables.

We sample and solve 50 instances of the problem for $N = 2, 3$ and $T = 2, 4, \ldots, 10$. Since $q_n = 100/N$ for all n and the splitting method facilitates the use of integer non-binary variables, we may substitute the N binary decision variables for each period by a single integer variable: $z_t(\boldsymbol{\zeta}_{1:t}) = \sum_{n=1}^{N} y_{nt}(\boldsymbol{\zeta}_{1:t})$ for all $t = 2, \ldots, T$, such that $0 \leq z_t(\boldsymbol{\zeta}_{1:t}) \leq N$ for all t. To see that this is possible, consider a fixed time period t and assume w.l.o.g. that $c_{y_1} \leq \ldots \leq c_{y_N}$. We know that if at the optimal solution exactly z_t of the variables $y_{1t}, \ldots, y_{Nt}$ have value 1, those will be the variables

Figure 3.8 – $N = 3$. The ordering cost of the fixed-size lots at time t at the optimal solution is a convex, piecewise linear function of z_t.



corresponding to the z_t smallest c_{y_n}'s:

$$\begin{aligned}\min_{y_{nt}} \left\{ \sum_{n=1}^{N} c_{y_n} q_n y_{nt} \,\middle|\, \sum_{n=1}^{N} y_{nt} = z_t \right\} &= \min_{y_{nt}} \left\{ \tfrac{100}{N} \sum_{n=1}^{N} c_{y_n} y_{nt} \,\middle|\, \sum_{n=1}^{N} y_{nt} = z_t \right\} \\ &= \tfrac{100}{N} \sum_{n=1}^{z_t} c_{y_n} \\ &= \tfrac{100}{N} \max_{n \in \{1,\ldots,N\}} \left\{ c_{y_n}(z_t - (n-1)) + \sum_{k=1}^{n-1} c_{y_k} \right\}. \end{aligned} \tag{3.20}$$

The last equality follows from the fact that when $c_{y_1} \leq \ldots \leq c_{y_N}$, then the sum of z_t smallest c_{y_n} is a convex piecewise linear function of z_t, which can be reformulated as a maximum over N linear functions of z_t, see Figure 3.8. For that reason, the obtained formulation can substitute the component $\sum_{n=1}^{N} c_{y_n} q_n y_{nt}$ in the objective function of (3.19) without losing the problem's convexity.

Since problem (3.19) involves fixed recourse only, we study also the impact of using linear decision rules for the continuous variables $x_t(\boldsymbol{\zeta}_{1:t})$. In such case we set $x_t(\boldsymbol{\zeta}_{1:t})$ to be an affine function of $\zeta_1, \ldots, \zeta_t$:

$$x_t(\boldsymbol{\zeta}_{1:t}) = \alpha_{t,0} + \sum_{j=1}^{t} \alpha_{t,j} \zeta_j, \quad \forall t = 1, \ldots, T-1,$$

where $\alpha_{t,j}$ are then treated as decision variables implemented in period t.

Each problem instance is solved in four ways: 1) applying static decisions to all variables 2) applying linear decision rules to the continuous variables and static decisions to the integer variables 3) applying only the splitting methodology to all variables 4) applying the splitting methodology to all variables, combined with linear decision

rules for the continuous decisions (the parameters $\alpha_{t,j}$ can also differ after splitting of the uncertainty set).

For each instance we conduct 4 splitting rounds. For splitting we use the worst-case scenario sets obtained using optimal KKT vectors from the robust counterpart of the LP relaxation of the problem (see Sections 3.2.3.2 and 3.3.3.2). In each splitting round we split all subsets $\mathcal{Z}_{r,s}$ for which $|\overline{\mathcal{Z}}_{r,s}| > 1$. Time periods t for the t-SHs are chosen according to the biggest variance of uncertain demands from subsequent periods with $q = 2$ (see Section 3.5.1). Splitting hyperplanes are constructed using Heuristic 3.1 (see Section 5.2). The scenario sets for the lower bound problems are constructed according to the 'simple heuristic' (see Section 3.5.3) with $k = 2$. For $T = 2, 4$ the lower bound scenario sets include also all vertices of the uncertainty set $\mathcal{Z}$. The after-splitting robust MILP problems are solved with Gurobi precision (the relative duality gap when the solver stops) equal to 0.1%. All problems were formulated using CVX package and solved with Gurobi solver on an Intel Core 2.66GHz computer.

Tables 3.4 and 3.5 give our results for $N = 2$ and $N = 3$, respectively. All methodologies offer substantial improvements in the objective value compared to the static robust solution. Also, combination of our splitting methodology with linear decision rules (S+LDR) gives a strong combined effect - the objective value improves significantly more than using any of the methods S or LDR separately - by as much as 64.82% for $N = 3, T = 10$, compared to 53.21% for LDR and 21.42% for S. For $T = 2$ the linear decision rules cannot bring any improvement because x_1 is a scalar. One can observe that for problems with larger T our methodology gives better objective improvements. Also, the relative optimality gaps decrease significantly in all cases, mostly due to improvements in the objective function. All problems have been solved fast, with the maximum mean time equal to 55.82s.

We compare now our results to those of Bertsimas and Georghiou (2015). The main difference between the methods lies in the fact that decision rules proposed by Bertsimas and Georghiou (2015) satisfy the problem's constraints with a high probability (99%), obtained using Hoeffding bounds, whereas our methodology ensures 100% robustness by design. Comparing the numbers from Tables 3.3 (column '$\mathcal{PB}_t(1)$ gap'), 3.4, and 3.5 (columns 'Final gap (%) - S+LDR'), one can see that our methodology performs worse in terms of the final optimality gap. For example, for $N = 2, T = 4$ our result is 39.16% compared to their 24.2% for $N = 2, T = 4$. This can be partly explained by the difference between the types of robustness, and also by different way of choosing the scenarios for the lower bounding problems. On the other hand, our method provides significantly faster computation times which, combined with full robustness, may be an appealing property. In particular, this is visible on larger instances, with our mean solution times being significantly lower, e.g., our 55.82s compared to 39141.5s for $N = 3, T = 10$.

Table 3.4 – Our results for the lot sizing problem for $N = 2$. LDR stands for the solution with linear decision rules for the continuous decision variables and static decisions for the integer variables, S stands for only our splitting methodology applied to all variables, S+LDR stands for a combination of set splitting with linear decision rules for the continuous variables. 'Objective improvement' is the decrease in the average worst-case objective value reduction, relative to the static robust solution. Optimality gaps are computed as in Table 3.3. 'Initial gap' is the optimality gap for the static robust solution and the lower bound obtained after the first splitting round. 'Final gap' is the optimality gap computed with the objective value and lower bound after the last splitting round. The asterisk indicates the fact that for $T = 2, 4$ the lower bound scenario sets include also all vertices of the uncertainty set $\mathcal{Z}$. All the static robust problems were solved in less than 2s.

	Objective improvement (%)			Initial gap (%)	Final gap (%)			Mean time (s)	
T	LDR	S	S+LDR		LDR	S	S+LDR	S	S+LDR
2	0	11.39	11.38	51.02∗	51.02∗	15.49∗	15.51∗	2.36	2.77
4	31.64	28.07	42.32	85.78∗	52.46∗	57.34∗	34.04∗	5.67	7.69
6	43.77	30.29	54.94	113.14	69.22	87.51	47.39	5.64	10.09
8	48.91	26.32	61.01	125.59	78.73	107.17	54.68	7.54	15.03
10	52.09	22.43	64.21	134.65	86.16	121.02	61.85	9.23	24.23

Table 3.5 – Our results for the lot sizing problem for $N = 3$. Terminology is the same as in Table 3.4.

	Objective improvement (%)			Initial gap (%)	Final gap (%)			Mean time (s)	
T	LDR	S	S+LDR		LDR	S	S+LDR	S	S+LDR
2	0	22.94	22.94	61.90∗	61.90∗	17.64∗	17.64∗	2.25	2.61
4	32.66	31.70	47.22	95.06∗	62.30∗	65.09∗	39.14∗	5.24	7.39
6	43.99	29.41	56.86	118.38	78.36	96.55	54.14	5.85	9.80
8	50.14	25.13	62.05	129.06	85.27	113.58	61.48	7.11	14.18
10	53.21	21.42	64.82	136.55	92.22	125.08	68.88	9.18	55.82

Table 3.6 – Lot sizing problem. Average-case performance of the solutions obtained using the three methodologies in comparison to the static robust solution. 'Average-case improvement' is the average reduction of the total cost, relative to the total costs obtained with the static solution for the given demand scenario.

	Average-case improvement (%)					
	$N = 2$			$N = 3$		
T	LDR	S	S+LDR	LDR	S	S+LDR
2	0.00	18.55	18.55	0.00	14.59	14.13
4	21.87	22.91	31.90	24.51	26.51	37.65
6	30.80	23.24	41.02	33.69	22.72	45.81
8	35.23	20.00	48.05	40.94	18.83	51.56
10	39.67	16.95	51.68	43.79	15.78	55.07

In addition to the worst-case results, for each solved instance we conduct a simulation study. In each of them we sample uniformly 500 demand scenario realizations $l \leq d_{\text{realized}} \leq u$ and compute the average total costs incurred by each of the four solutions. Table 3.9 gives the results on average-case improvements relative to the static robust solution. The table shows that our method not only offers substantial improvements on the worst-case basis, but also in terms of the average-case total cost, in particular when combined with the linear decision rules for the continuous variables.

To sum up the results of this numerical example, the main benefits of our approach have been: 1) fast computation time even for large problems, corresponding to the number of splitting rounds (the more splitting rounds, the better the improvement in the objective, but also the longer computation time), 2) substantial improvements in the objective function value, 3) robustness to the entire uncertainty set after each splitting round.

3.6.3 Route planning

We consider another numerical example from Hanasusanto et al. (2015), the route planning problem, where the uncertainty occurs only in the objective function. On this example, we shall see that our methodology depends heavily on having multiple uncertain constraints that give rise to different worst-case scenarios for the uncertain parameter.

The problem at hand is a shortest path problem that is defined on a directed, arc-weighted graph $G = (V, A, \boldsymbol{w})$ with nodes $V = \{1, \ldots, N\}$, arcs $A \subseteq V \times V$, and weights $w_{ij}(\xi) \in \mathbb{R}_+$, $(i, j) \in A$. We assume that the arc weights $w_{i,j}$ are functions of an uncertain parameter vector $\boldsymbol{\zeta}$ that is only known to reside in an uncertainty set $\mathcal{Z}$. The goal is to determine the shortest worst-case path from a start node $b \in V$ to a terminal node $e \in V$, $b \neq e$, before the value of $\boldsymbol{\zeta}$ is known. Hanasusanto et al.

(2015) consider the number of possible paths to be fixed and equal to K.

In our setting, we begin with the following robust problem, equivalent to having a single path. The binary variable x_{ij} is equal to 1 if arc (i, j) is a part of the path from b to e:

$$\begin{aligned} \min_{z,\boldsymbol{x}} \quad & z \\ \text{s.t.} \quad & \sum_{(i,j)\in A} w_{ij}(\boldsymbol{\zeta}) x_{ij} \leq z, \quad \forall \boldsymbol{\zeta} \in \mathcal{Z} \\ & x_{ij} \in \{0,1\}, \quad (i,j) \in A \\ & \sum_{(j,l)\in A} x_{jl} \geq \sum_{(i,j)\in A} x_{ij} + \mathbb{I}(j=b) - \mathbb{I}(j=e), \ \forall j = 1,\ldots,N, \end{aligned} \tag{3.21}$$

where $\mathbb{I}(\cdot)$ is the indicator function. In our method the set $\mathcal{Z}$ is split into subsets, to each of which a separate route vector $\boldsymbol{x}^{(r,s)}$ shall correspond. That is, when $\boldsymbol{\zeta} \in \mathcal{Z}_{r,s}$, then the corresponding path vector $\boldsymbol{x}^{(r,s)}$ is chosen. Such a problem has the property that finding the optimal path for each $\mathcal{Z}_{r,s}$ can be solved as a separate optimization problem, solving thus N_r smaller problems instead of one large problem.

As visible in (3.21), there is only one uncertain constraint in this problem. For that reason, we use in this case the sets $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r,s)})$ obtained by searching for the critical scenarios based on the primal solution $\boldsymbol{x}^{(r)}$. However, as there is only one uncertain constraint, solving problem (3.17) for this constraint results in only one critical scenario $\overline{\boldsymbol{\zeta}}$. However, we need at least two distinct scenarios to be divided with a splitting hyperplane.

We propose, for a given subset $\mathcal{Z}_{r,s}$ with the corresponding optimal solution vector $\boldsymbol{x}^{(r,s)}$, to choose the second member of $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r,s)})$ according to the following procedure:

1. Find an alternative route from b to e that uses at most $\lfloor \theta \mathbf{1}^T \overline{\boldsymbol{x}}^{(r,s)} \rfloor$ arcs from the path corresponding to $\overline{\boldsymbol{x}}^{(r,s)}$, where $0 \leq \theta \leq 1$ denotes the fraction of the arcs from the 'old' path allowed to use. Denote the new alternative vector by $\widetilde{\boldsymbol{x}}^{(r,s)}$.

2. Find the worst-case scenario $\widetilde{\boldsymbol{\zeta}}$ corresponding to $\widetilde{\boldsymbol{x}}^{(r,s)}$ and add it to $\overline{\mathcal{Z}}_{r,s}(\overline{\boldsymbol{x}}^{(r,s)})$. If no feasible $\widetilde{\boldsymbol{x}}^{(r,s)}$ exists, take $\widetilde{\boldsymbol{\zeta}}$ to be the scenario that *minimizes* the uncertain distance corresponding to path $\overline{\boldsymbol{x}}^{(r,s)}$ (in contrast to the worst-case scenarios, which *maximize* the uncertain distance corresponding to path $\overline{\boldsymbol{x}}^{(r,s)}$).

We consider three values for θ: 0, 0.5, and 0.9. and adopt the same data setting as Hanasusanto et al. (2015). Table 3.7 presents the improvement results obtained by Hanasusanto et al. (2015). We sample 40 problem instances and, for each instance, we allow 90s for the subsequent splitting rounds, with at most 10s for each optimization

Table 3.7 – Results of Hanasusanto et al. (2015) for the route planning problem. K is the number of time-2 decisions, B denotes the size of the uncertainty set, and N is the number of nodes. The columns are (1) - percentage of instances solved to optimality within 2h (%), (2) - average solution time of the instances solved within 2h (in seconds), (3) - average objective improvements (%).

	$B=3$									$B=6$								
	$K=2$			$K=3$			$K=4$			$K=2$			$K=3$			$K=4$		
N	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
20	100	8	8.31	97	463	10.26	51	1103	10.70	100	7	6.07	97	428	10.23	55	795	11.79
25	99	168	9.49	31	1273	12.06	6	2851	12.81	99	197	8.39	38	1771	13.36	7	3208	15.48
30	69	1131	9.51	6	1563	12.79	0	-	13.94	67	1372	9.06	6	1537	14.61	0	-	17.38
35	17	2335	9.97	0	-	13.70	0	-	15.20	16	2819	9.90	0	-	15.86	0	-	19.20
40	6	2949	9.94	0	-	13.74	0	-	15.28	5	2888	10.36	0	-	16.47	0	-	19.84
45	0	-	9.46	0	-	13.39	0	-	15.19	0	-	10.67	0	-	16.78	0	-	20.32
50	0	-	9.38	0	-	13.31	0	-	15.14	0	-	10.71	0	-	16.73	0	-	20.30

problem to solve. Afterwards, we allow a solution time of 60s for each problem. For splitting the uncertainty subsets we use Heuristic 3.1.

Table 3.8 presents the results on the improvement in the objective function value. One can see that the method of Hanasusanto et al. (2015) performs significantly better than our approach, with the difference growing with N and the number K of possible time-2 decisions they use. For example, whereas for problems with $B = 3$ and $N = 20$ we obtain improvement of 5.50% compared to their 8.31% for $K = 2$ and 10.70% for $K = 4$, for problems with $B = 6$ and $N = 50$ we get 6.70% and they obtain 20.30% for $K = 4$. Additionally, one can see that our approach performs best for $\theta = 0$, decreasing with larger values of θ.

The big difference between the performances of our methodology and the one of Hanasusanto et al. (2015) is most likely due to the fact that their methodology optimizes the fixed number of K of decisions. This is implicitly equivalent to optimizing the division of the uncertainty set into K regions corresponding to K possible decisions (each possible time 2 decision has its 'share' of the uncertainty set on which it is at least as good as the other decision). On the other hand, in our methodology the splits are chosen in a relatively simple manner, by means of heuristics, which in this particular case do not perform very well if it is even not known exactly which scenarios should be separated by the splitting hyperplane. The impact of the difference of allowed solution time - 7200s by Hanasusanto et al. (2015) and average time of 184s in our case - is not expected to be substantial as Hanasusanto et al. (2015) report that in their case the terminal solution was attained typically after 60s.

Remark 3.1 *The results of the route planning experiment leads to a remark that in fact, the methodology of Hanasusanto et al. (2015) could be used to construct 'best' splitting hyperplanes in multiperiod problems with only binary decision variables. In*

Table 3.8 – Our results for the route planning problem. Average solution time over all instances was equal to 184s.

	Improvement (%)					
	$\theta = 0$		$\theta = 0.5$		$\theta = 0.9$	
N	$B = 3$	$B = 6$	$B = 3$	$B = 6$	$B = 3$	$B = 6$
20	5.50	3.91	4.29	3.88	1.65	2.80
25	6.15	5.01	4.44	4.27	2.63	3.18
30	6.26	5.12	4.18	5.44	1.78	2.73
35	6.34	6.01	4.39	5.73	2.02	3.21
40	5.68	6.90	4.09	5.89	1.07	3.65
45	5.28	7.96	3.76	5.35	1.79	2.38
50	5.45	6.70	2.79	5.18	0.61	2.85

such a case, in case of splitting Heuristic 3.3 (see Section 3.5.2), one would no longer keep the normal vector $\boldsymbol{g}$ of the hyperplane fixed, but would optimize it jointly with h. This, however, would only be possible for problems where the methodology of Hanasusanto et al. (2015) is applicable, i.e., to problems with adjustable binary variables.

3.6.4 Inventory management - continuous decisions

From the results of Section 3.6.2 it is clear that a substantial gain of the splitting methodology is in splitting w.r.t. the continuous variables alone. For that reason, in this section we consider an experiment in which only continuous decision variables are used.

The problem studied is the same as in Section 3.6.2, but without the integer decisions y_{nt}. Since the integer decisions y_{nt} have played a substantial role in feasibility of the solutions in Section 3.6.2 (meeting the demand, since the inventory level is assumed to be nonnegative) simply leaving them out would lead to many of the problem instances being infeasible. For that reason, we set the new upper bounds on the total orders equal to $\overline{x}_{\text{tot},t} := 3\overline{x}_{\text{tot},t}$ for all time periods t, i.e., we set them to be three times larger.

Also, since problems with continuous variables only are solved fast, we set the maximum number of splits to 10 and $q = 1$ (see Section 3.5.1). We solve the problem for $T \in \{6, 8, 10\}$. Out of 50 problem instances, 28 turned out to be feasible.

There results are presented in Table 3.9 and there are two interesting phenomena that can be observed. First of all, in column 7 it can be observed that there is no optimality gap for the solutions S+LDR. However, looking at the columns 1 and 3, we can see that in fact the solutions LDR and S+LDR provided the same objective

Table 3.9 – Results for the inventory experiment with continuous variables only. Terminology is the same as in Table 3.4.

	Objective improvement (%)			Initial gap (%)	Final gap (%)		
T	LDR	S	S+LDR		LDR	S	S+LDR
6	63.37	34.97	63.37	118.72	35.14	71.96	0
8	71.46	30.66	71.46	139.11	47.36	104.11	0
10	76.23	26.61	76.23	151.10	56.19	124.77	0

function values, i.e., that linear decision rules are optimal for the problem at hand and no splitting is needed if they are used. As we can see, however, the fact that they are optimal is only visible from the lower bounds obtained for S+LDR, i.e., after several splitting rounds where many uncertainty scenarios have been added to the scenario set and the lower-bound problem has been used. Thus, despite the fact that with linear decision rules the splitting cannot improve the objective, it can help to obtain a scenario set giving the best possible lower bounds.

Second, to compare the gain obtained by splitting itself, it is instructive to compare the objective improvements for solutions S, e.g., in Table 3.4 with the objective improvements in column S in Table 3.9. The objective improvements are 29.41, 25.13, 21.42 vs 34.97, 30.66, 26.61, respectively. Therefore, one can say that compared to Section 3.6.2, there is more gain here due to the splitting. However, it might not be reasonable to generalize this conclusion to say that in problems with continuous variables only the gain is always larger.

3.7 Conclusions

In this paper we have introduced the method of iterative splitting of the uncertainty set for multi-period robust mixed-integer linear optimization problems. We have provided theory on how to determine efficiently which scenarios of the uncertain parameter are more important to be separated than others and how to obtain lower bounds for the adjustable worst-case value. Based on these theoretical results, we have proposed several heuristics for each part of the method.

Our approach can be used to a variety of problems. In particular, this applies to problems with a non-fixed recourse and adjustable integer variables (also non-binary), where implementation of other decision rules may not be possible or may involve large computational effort. For adjustable continuous variables in the non-fixed recourse setting, our method bypasses the challenge of dealing with interactions of uncertain parameters, as would be the case with linear or polynomial decision rules.

For fixed recourse problems the splitting method can be combined with other decision rules, such as linear decision rules, allowing them to take different forms over differ-

ent parts of the uncertainty set. The second numerical experiment reveals that such a combination gives a strong joint effect. Our iterative method guarantees robustness of the decisions to the entire uncertainty set after each of the splitting rounds. Thus, depending on time constraints, the decision maker can set how many splitting rounds to conduct, with each additional round costing additional effort but bringing potentially extra improvements in the objective value.

Numerical experiments conducted on problems from Bertsimas and Georghiou (2015) and Hanasusanto et al. (2015) have shown our methodology to perform well on problems involving non-constraint-wise uncertainty. In both cases was our method outperformed on small problem instances. However, as the problems grow, our methodology was giving comparable results after only a fraction of the computation time of other authors.

We give now potential directions for further research. First, more theoretical results can be obtained regarding the choice of best splits of the uncertainty sets, and in particular, the 'best' distribution of the splits in time. Secondly, it is important to obtain better lower bound values, possibly by combining our method with results of other authors, e.g., Kuhn et al. (2011). Last, it is interesting to investigate whether our method, combined with the results of Ben-Tal et al. (2015), can be used efficiently in multistage nonlinear robust optimization problems.

Bibliography

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

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

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

A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.

D. Bertsimas and C. Caramanis. Adaptability via sampling. In *46th IEEE Conference on Decision and Control*, pages 4717–4722. IEEE, 2007.

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. Binary decision rules for multistage adaptive mixed-integer optimization. http://www.optimization-online.org/DB_FILE/2014/08/4510.pdf, 2014.

D. Bertsimas and A. Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, 63(3):610–627, 2015.

D. Bertsimas and V. Goyal. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, 35(2):284–305, 2010.

D. Bertsimas, D. A. Iancu, and P. A. Parrilo. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

D. Bertsimas, V. Goyal, and X. A. Sun. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research*, 36(1):24–54, 2011a.

D. Bertsimas, D. A. Iancu, and P. A. Parrilo. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12): 2809–2824, 2011b.

D. Bertsimas, V. Goyal, and B. Y. Lu. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming*, 150(2):281–319, 2014.

J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

J. R. Birge and J.-B. Wets. Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Mathematical Programming Study*, 27:54–102, 1986.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

X. Chen and Y. Zhang. Uncertain linear programs: extended affinely adjustable robust counterparts. *Operations Research*, 57(6)(6):1469–1482, 2009.

X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

K. Frauendorfer and P. Kall. A solution method for SLP recourse problems with arbitrary multivariate distributions - the independent case. *Problems in Control and Information Theory*, 17:177–205, 1988.

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.

V. Goyal and B. Y. Lu. On the adaptivity gap in two-stage robust linear optimization under uncertain constraints. http://www.optimization-online.org/DB_FILE/2014/12/4706.pdf, 2014.

M. J. Hadjiyiannis, P. J. Goulart, and D. Kuhn. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *IEEE Conference on Decision and Control and European Control Conference*. IEEE, 2011.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage robust binary programming. *Operations Research*, 63(4):877–891, 2015.

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

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.

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

CHAPTER 4

Robust optimization with ambiguous stochastic constraints under mean and dispersion information

4.1 Introduction

Consider an optimization problem with a constraint

$$f(\boldsymbol{x}, \boldsymbol{z}) \leq 0,$$

where $\boldsymbol{x} \in \mathbb{R}^{n_{\boldsymbol{x}}}$ is the decision vector, $\boldsymbol{z} \in \mathbb{R}^{n_{\boldsymbol{z}}}$ is an uncertain parameter vector, and $f(\cdot, \boldsymbol{z})$ is assumed to be convex for all $\boldsymbol{z}$. There are three principal ways to address such constraints. One of them is *Robust Optimization*. In this approach, $\mathcal{U}$ is a user-provided convex compact uncertainty set and the constraint is to hold for all $\boldsymbol{z} \in \mathcal{U}$, i.e., $\boldsymbol{x}$ is robust feasible if:

$$\sup_{\boldsymbol{z} \in \mathcal{U}} f(\boldsymbol{x}, \boldsymbol{z}) \leq 0. \tag{4.1}$$

The key issue in this approach is to reformulate (4.1) to an equivalent, computationally tractable form (Ben-Tal and Nemirovski (1998), Ben-Tal et al. (2009, 2015)).

In the other approaches, which go under the name of *Distributionally Robust Optimization* (DRO), $\boldsymbol{z}$ is a *random* parameter vector whose distribution $\mathbb{P}_{\boldsymbol{z}}$ belongs to a set $\mathcal{P}$ (the so-called *ambiguity set*). A typical example for $\mathcal{P}$ is a set of all distributions with given values of the first two moments. In such a setting, there are two principal constraint types: the *worst-case expected feasibility constraints*:

$$\sup_{\mathbb{P}_{\boldsymbol{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\boldsymbol{z}}} f(\boldsymbol{x}, \boldsymbol{z}) \leq 0, \tag{4.2}$$

and *chance constraints*:

$$\sup_{\mathbb{P}_{\boldsymbol{z}} \in \mathcal{P}} \mathbb{P}_{\boldsymbol{z}} \left(f(\boldsymbol{x}, \boldsymbol{z}) > 0 \right) \leq \epsilon. \tag{4.3}$$

For constraint (4.2) the key challenge is, for a given ambiguity set $\mathcal{P}$, to obtain a tractable exact form of the worst-case expectation, or a good upper bound. Constraint (4.2) is also used in the construction of safe approximation of the ambiguous chance constraint (4.3), where by a *safe approximation* is meant a system $\mathcal{S}$ of computationally tractable constraints, such that $\boldsymbol{x}$ feasible for $\mathcal{S}$ is also feasible for constraint (4.3).

In this paper, we consider problems with ambiguity sets consisting of distributions having given mean-dispersion measures. The literature of this type of problem started with the paper by Scarf (1958). Under mean-variance information, Scarf derived the exact worst-case expectation formula for a single-variable piecewise linear objective function used in the newsvendor problem. Later, his result has been extended to more elaborate cases of inventory and newsvendor problems by, e.g., Gallego (1992), Gallego et al. (2001), and Perakis and Roels (2008). In a paper by Popescu (2007), it has been proved that for a wide class of increasing concave utility functions the problem of maximizing the worst-case expected utility under mean-variance distributional information reduces to solving a parametric quadratic optimization problem.

In a broader context, the idea of constructing an approximation of the worst-case expectation of a given function by a discrete distribution falls into the category of bounding strategies based on distributional approximation, see Edirisinghe (2011) who provide a broad overview of results obtained in this field. Rogosinsky (1958) and Karr (1983) show that the worst-case probability distributions corresponding to the moment problems are discrete, with a number of points corresponding to the number of moment conditions. Shapiro and Kleywegt (2002) develop a duality theory for stochastic programs where the saddle points are also vectors of discrete probabilities. Dupačová (1966) and Gassmann and Ziemba (1986) give convex upper bounds on the expectation of a convex function under first-moment conditions over a polyhedral support, based on the dual of the related moment problem. Birge and Wets (1987) and Edirisinghe and Ziemba (1994b) extend this approach to distributions with unbounded support. Dulá (1992) provide a bound for the expectation of a simplicial function of a random vector using first moments and the sum of all variances. His approach is extended by Kall (1991) demonstrating that the related moment problems can be solved using nonsmooth optimization problems with linear constraints. Other notable works in this field include Frauendorfer (1988), Edirisinghe and Ziemba (1992), and Edirisinghe and Ziemba (1994a). For a general discussion we refer the reader to Edirisinghe (2011) and references therein.

Despite numerous works, to the best of our knowledge, no *closed-form tight upper bounds* are known on the expectations of *general convex functions* under mean-variance information. Surprisingly, already in 1972 a result of Ben-Tal and Hochman (1972) was available, providing exact upper and lower bounds on the expectation of

a general convex $f(\boldsymbol{x}, \cdot)$ for the case where $\mathcal{P}$ consists of all distributions of componentwise independent $\boldsymbol{z}$ with known supports, means, but with another dispersion measure: the *mean absolute deviation from the mean (MAD)*.

The information about the supports, means, and MADs of the z_i's, can easily be obtained from past data, making the method suitable for data-driven settings. Moreover, the MAD has several desirable properties from an application's point of view, for example, its suitability to situations when the deviations of z_i are small. Some properties of the MAD and exact formulas for its value for several known distributions are given in Appendix 4.D. Several practical advantages of using the MAD as a dispersion measure are given in the paper by El Amir (2012) and references therein.

Our contributions can be summarized as follows:

- We propose a new method of optimizing the *exact* worst-case-expected performance under mean-MAD information in problems involving constraints (4.2) with both convex and concave $f(\boldsymbol{x}, \cdot)$, or a mixture of these.
- We derive new safe tractable approximations of chance constraint (4.3) under mean-MAD information. These results apply to the case of independent random variables z_i.
- In problems where the random variables are linearly aggregated, i.e.

  $$f(\boldsymbol{x}, \boldsymbol{a}^T\boldsymbol{z}) \text{ or } f(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z}),$$

  we derive upper bounds which do not require the independence of the random variables and which are computationally tractable.
- The above results are used to treat problems in which convexity in the uncertain parameter (usually an intractable case in classical RO) appears. This occurs, for example as a result of applying linear decision rules or when the uncertainty is due to implementation error.
- Moreover, in case of existence of multiple RO-optimal solutions, we show that the proposed approach can be used as a second-stage method of improving the *average* performance of the RO solutions.
- Our numerical study shows that minimization of $\sup_{\mathbb{P}_{\boldsymbol{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\boldsymbol{z}}} f(\boldsymbol{x}, \boldsymbol{z})$ over $\boldsymbol{x}$ using mean-MAD information can also lead to a downward shift of $\inf_{\mathbb{P}_{\boldsymbol{z}} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}_{\boldsymbol{z}}} f(\boldsymbol{x}, \boldsymbol{z})$, compared to the solution obtained by classical RO.

We mention that there are alternative ways of specifying the set $\mathcal{P}$, for example, as sets of distributions deviating from a known distribution according to a certain

distance measure (see for example Ben-Tal et al. (2013)). For a broad overview of types of ambiguity sets we refer the reader to Postek et al. (2015) and Hanasusanto et al. (2015). There are cases of such alternative settings for which exact reformulations are possible both for the expected feasibility constraints and for the chance constraints. For the first case, examples of such settings are given in Ben-Tal et al. (2013), Wiesemann et al. (2014) and Esfahani and Kuhn (2015). Settings in which exact reformulations are possible for individual chance constraints are given, for example, in Calafiore and El Ghaoui (2006) and Jiang and Guan (2016).

Approximation results or efficient solution methods when the components of the random vector $\boldsymbol{z}$ are not independent, are obtained for limited classes of function $f(\boldsymbol{x}, \boldsymbol{z})$ in Delage and Ye (2010), Goh and Sim (2010), and Zymler et al. (2013). Chen et al. (2007) propose to use so-called forward and backward deviations as characteristics of the moment generating functions of random variables to approximate chance constraints.

Wiesemann et al. (2014) have recently introduced a class of quite general ambiguity sets for which they derive computationally tractable counterparts of (4.2) for specific cases of $f(\boldsymbol{x}, \cdot)$. However, in their framework the components of $\boldsymbol{z}$ are unrestricted in their dependence, and taking their independence into account is not straightforward. In Appendix 4.C, we illustrate the marked difference between theirs and our robust counterparts when $f(\boldsymbol{x}, \boldsymbol{z}) = \exp(\boldsymbol{x}^T \boldsymbol{z})$ where, without the assumption of independence, one has to reformulate a robust constraint that is *strictly convex* in the uncertain parameter. In Section 4.4.4 we provide a numerical comparison of their and our method on an example where both approaches can be applied without the independence assumption.

The remainder of the paper is structured as follows. In Section 4.2, we describe the mean-MAD results of Ben-Tal and Hochman (1972), providing statistical background on estimation of the relevant parameters. In Section 4.3, we show how the mean-MAD results can be used to optimization problems involving stochastic constraints (4.2), including numerical examples. In Section 4.4 we outline the result for the case of linearly aggregated random variable. Section 4.5 includes new results on safe tractable approximations of chance constraints (4.3), illustrated also with a numerical study. Section 4.6 concludes the paper.

4.2 Expectation of a convex function of a random variable

4.2.1 Introduction

In this section we introduce the results of Ben-Tal and Hochman (1972) on exact upper and lower bounds on the expected value of a convex function of a componen-

twise independent $\boldsymbol{z} = (z_1, \ldots, z_{n_{\boldsymbol{z}}})^T$. From now on we drop the subscript $\boldsymbol{z}$ from $\mathbb{P}_{\boldsymbol{z}}$ and the probability distribution applies to $\boldsymbol{z}$. The pieces of partial distributional information on z_i's constituting the ambiguity sets in Ben-Tal and Hochman (1972) are:

1. support including intervals: $\text{supp}(z_i) \subseteq [a_i, b_i]$, where $-\infty < a_i \leq b_i < \infty, i = 1, \ldots, n_{\boldsymbol{z}}$. Ben-Tal and Hochman (1972) show also that their bounds hold in cases where $a_i = -\infty$ and/or $b_i = +\infty$. We illustrate this in Remark 4.3. In the remainder of the paper, however, we concentrate on the bounded case, with RO applications in mind,

2. means: $\mathbb{E}_{\mathbb{P}}(z_i) = \mu_i$,

3. mean absolute deviations from the means (MAD): $\mathbb{E}_{\mathbb{P}}|z_i - \mu_i| = d_i$. The MAD is known to satisfy the bound (Ben-Tal and Hochman (1972), Lemma 1):

$$0 \leq d_i \leq d_{i,\max} = \frac{2(b_i - \mu_i)(\mu_i - a_i)}{(b_i - a_i)}, \quad i = 1, \ldots, n_{\boldsymbol{z}}, \tag{4.4}$$

4. probabilities of z_i's being greater than or equal to μ_i: $\mathbb{P}(z_i \geq \mu_i) = \beta_i$. For example, in the case of continuous symmetric distribution of z_i we know that $\beta_i = 0.5$. This quantity is known to satisfy the bounds:

$$\frac{d_i}{2(b_i - \mu_i)} = \underline{\beta}_i \leq \beta_i \leq \overline{\beta}_i = 1 - \frac{d_i}{2(\mu_i - a_i)}, \quad i = 1, \ldots, n_{\boldsymbol{z}}. \tag{4.5}$$

Using these building blocks, we define two types of ambiguity set $\mathcal{P}$:

- the (μ, d) *ambiguity set*, consisting of the distributions with known (i), (ii), and (iii) for each z_i:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P} : \quad \text{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_{\mathbb{P}}(z_i) = \mu, \quad \mathbb{E}_{\mathbb{P}}\,|z_i - \mu_i| = d_i, \quad \forall i, \quad z_i \perp\!\!\!\perp z_j, \quad \forall i \neq j\}\,, \tag{4.6}$$

  where $z_i \perp\!\!\!\perp z_j$ denotes the stochastic independence of components z_i and z_j,

- the (μ, d, β) *ambiguity set*, consisting of the distributions with known (i), (ii), (iii), and (iv) for each z_i:

$$\mathcal{P}_{(\mu,d,\beta)} = \left\{\mathbb{P} : \quad \mathbb{P} \in \mathcal{P}_{(\mu,d)}, \quad \mathbb{P}(z_i \geq \mu_i) = \beta_i, \quad \forall i\right\}. \tag{4.7}$$

In the following, we present the results of Ben-Tal and Hochman (1972) on $\max\limits_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z})$, $\max\limits_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z})$ and $\min\limits_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z})$, $\min\limits_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z})$, where $f : \mathbb{R}^{n_{\boldsymbol{z}}} \to \mathbb{R}$ is convex. We note that in the case of concave $f(\cdot)$ the upper bounds become lower bounds and vice versa.

4.2.2 One-dimensional z

We begin with the simpler and more illustrative case of one-dimensional random variable z. For that reason, we drop the subscript i.

Upper bounds. Ben-Tal and Hochman (1972) shows that:

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = p_1 f(a) + p_2 f(\mu) + p_3 f(b), \tag{4.8}$$

where:

$$p_1 = \frac{d}{2(\mu - a)}, \quad p_2 = 1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}, \quad p_3 = \frac{d}{2(b - \mu)}. \tag{4.9}$$

Hence, the worst-case distribution is a three-point distribution on $\{a, \mu, b\}$. The same bound holds for the (μ, d, β) ambiguity set.

Remark 4.1 *A special case of (4.9) is the upper bound on $f(z)$ when only the interval $[a, b]$ and the mean μ are known. Such a bound is known as the Edmundson-Madansky bound (Edmundson 1956, Madansky 1959):*

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu)}} \mathbb{E}_{\mathbb{P}} f(z) = \frac{b-\mu}{b-a} f(a) + \frac{\mu - a}{b-a} f(b) \text{ where } \mathcal{P}_{(\mu)} = \{\mathbb{P} : \quad supp(z) \subseteq [a, b], \quad \mathbb{E}_{\mathbb{P}} z = \mu\}. \tag{4.10}$$

Indeed, inserting the biggest possible value of MAD (see (4.4)) equal to $d_{max} = 2(b - \mu)(\mu - a)/(b - a)$ into (4.9) yields the probability of outcome μ equal to 0. ■

Lower bounds. To obtain a closed-form lower bound on $\mathbb{E}_{\mathbb{P}} f(z)$, additional information is needed in the form of the parameter β. Then, it holds that:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(z) = \beta f\left(\mu + \frac{d}{2\beta}\right) + (1 - \beta) f\left(\mu - \frac{d}{2(1-\beta)}\right). \tag{4.11}$$

In case β is not known, Ben-Tal and Hochman (1972) shows that:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = \min_{\underline{\beta}\le\beta\le\overline{\beta}} \left\{\beta f\left(\mu + \frac{d}{2\beta}\right) + (1 - \beta) f\left(\mu - \frac{d}{2(1-\beta)}\right)\right\}, \tag{4.12}$$

where the minimization over β is a convex problem in β and for a strictly convex $f(\cdot)$ there is a unique optimal solution.

Remark 4.2 *In case of no knowledge about d, the lower bound is obtained at $d^* = 0$, which corresponds to the well-known Jensen bound (Jensen 1906).* ■

Remark 4.3 *In case where $a = -\infty$ and/or $b = +\infty$, bounds can still be obtained under additional conditions, namely that the limits $\lim_{t\to\pm\infty} f(t)/t$ exist and are finite, with the '+' corresponding to $b = +\infty$, and the '−' corresponding to $a = -\infty$. We illustrate this on the example $a \in \mathbb{R}, b = +\infty$. Assume that $\lim_{t\to+\infty} f(t)/t = \gamma$. We then have:*

$$\begin{aligned}\max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(z) = \max_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(z) \quad &= \lim_{b\to\infty} \left\{ \tfrac{d}{2(\mu-a)} f(a) + \left(1 - \tfrac{d}{2(\mu-a)} - \tfrac{d}{2(b-\mu)}\right) f(\mu) + \tfrac{d}{2(b-\mu)} f(b) \right\} \\ &= \tfrac{d}{2(\mu-a)} f(a) + \left(1 - \tfrac{d}{2(\mu-a)}\right) f(\mu) + \tfrac{d}{2}\gamma,\end{aligned}$$

and for the lower bound we have:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} f(z) = \frac{d}{2}\gamma + f\left(\mu - \frac{d}{2}\right).$$

The lower bound for the (μ, d, β) ambiguity set is the same as (4.11). ■

4.2.3 Multidimensional $\boldsymbol{z}$

Upper bounds. For $n_{\boldsymbol{z}} > 1$, the worst-case probability distribution under (μ, d) information is a componentwise counterpart of (4.9):

$$p_1^i = \frac{d_i}{2(\mu_i - a_i)}, \quad p_2^i = 1 - \frac{d_i}{2(\mu_i - a_i)} - \frac{d_i}{2(b_i - \mu_i)}, \quad p_3^i = \frac{d_i}{2(b_i - \mu_i)}, \quad i = 1, \ldots, n_{\boldsymbol{z}}. \tag{4.13}$$

The worst-case expectation of $f(\boldsymbol{z})$ is obtained by applying the bound (4.8) for each z_i, i.e., by enumerating over all $3^{n_{\boldsymbol{z}}}$ permutations of outcomes a_i, μ_i, b_i of components z_i. It holds then that (Ben-Tal and Hochman 1972):

$$\max_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z}) = \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^{n_{\boldsymbol{z}}}} \prod_{i=1}^{n_{\boldsymbol{z}}} p_{\alpha_i}^i f(\tau_{\alpha_1}^1, \ldots, \tau_{\alpha_{n_{\boldsymbol{z}}}}^{n_{\boldsymbol{z}}}), \tag{4.14}$$

where

$$\tau_1^i = a_i, \quad \tau_2^i = \mu_i, \quad \tau_3^i = b_i \quad \text{for} \quad i = 1, \ldots, n_{\boldsymbol{z}}. \tag{4.15}$$

Again, the same upper bound holds for the (μ, d, β) ambiguity set.

Lower bounds. Similar to the one-dimensional case, the closed-form lower bound under (μ, d) information requires known $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n_{\boldsymbol{z}}})^T$:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z}) = \sum_{\boldsymbol{\alpha}\in\{1,2\}^{n_{\boldsymbol{z}}}} \prod_{i=1}^{n_{\boldsymbol{z}}} q_{\alpha_i}^i f(\upsilon_{\alpha_1}^1, \ldots, \upsilon_{\alpha_{n_{\boldsymbol{z}}}}^{n_{\boldsymbol{z}}}), \tag{4.16}$$

where $\underline{\boldsymbol{\beta}} = (\underline{\beta}_1, \ldots, \underline{\beta}_{n_{\boldsymbol{z}}})^T, \overline{\boldsymbol{\beta}} = (\overline{\beta}_1, \ldots, \overline{\beta}_{n_{\boldsymbol{z}}})^T$ and

$$q_1^i = \beta_i, \quad q_2^i = 1 - \beta_i, \quad \upsilon_1^i = \mu_i + d_i/2\beta_i, \quad \upsilon_2^i = \mu_i - d_i/2(1 - \beta_i). \tag{4.17}$$

If $\boldsymbol{\beta}$ is unknown, the bound is obtained by minimization:

$$\min_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{z}) = \inf_{\underline{\boldsymbol{\beta}}\leq\boldsymbol{\beta}\leq\overline{\boldsymbol{\beta}}} \sum_{\boldsymbol{\alpha}\in\{1,2\}^{n_{\boldsymbol{z}}}} \prod_{i=1}^{n_{\boldsymbol{z}}} q^i_{\alpha_i} f(\upsilon^1_{\alpha_1},\ldots,\upsilon^{n_{\boldsymbol{z}}}_{\alpha_{n_{\boldsymbol{z}}}}). \tag{4.18}$$

In the multidimensional case, minimization over $\boldsymbol{\beta}$ is a nonconvex problem - it is only convex in β_i when other $\beta_j, j\neq i$ are fixed. A statistical procedure for estimating the parameters μ, d, and β is provided in Appendix 4.A.

4.3 Robust counterparts of expected feasibility constraints

4.3.1 Reformulations

In this section we demonstrate how the results of Ben-Tal and Hochman (1972) can be used to solve problems

$$\text{Val} = \min_{\boldsymbol{x}} \max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{x},\boldsymbol{z}), \tag{4.19}$$

where $f(\cdot,\boldsymbol{z})$ is convex. When $f(\boldsymbol{x},\cdot)$ is convex and $\mathcal{P}=\mathcal{P}_{(\mu,d)}$, the exact solution of the inner problem is given due to the Ben-Tal and Hochman (1972) upper bound:

$$g_U(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^{n_{\boldsymbol{z}}}} \prod_{i=1}^{n_{\boldsymbol{z}}} p^i_{\alpha_i} f(\boldsymbol{x},\tau^1_{\alpha_1},\ldots,\tau^{n_{\boldsymbol{z}}}_{\alpha_{n_{\boldsymbol{z}}}}), \tag{4.20}$$

with $p^i_{\alpha_i}, \tau^i_{\alpha_i}$ defined as in (4.13) and (4.15). As we can see, $g_U(\cdot)$ in (4.20) inherits the convexity in $\boldsymbol{x}$ from $f(\cdot,\boldsymbol{z})$ and its functional form depends only on the form of $f(.,\boldsymbol{z})$.

When $f(\boldsymbol{x},\cdot)$ is concave and $\mathcal{P}=\mathcal{P}_{(\mu,d,\beta)}$, the exact solution of the inner problem is given due to the Ben-Tal and Hochman (1972) lower bound:

$$g_L(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\{1,2\}^{n_{\boldsymbol{z}}}} \prod_{i=1}^{n_{\boldsymbol{z}}} q^i_{\alpha_i} f(\boldsymbol{x},\upsilon^1_{\alpha_1},\ldots,\upsilon^{n_{\boldsymbol{z}}}_{\alpha_{n_{\boldsymbol{z}}}}), \tag{4.21}$$

with $q^i_{\alpha_i}, \upsilon^i_{\alpha_i}$ defined by (4.17).

For the case of convexity (concavity) of $f(\boldsymbol{x},\cdot)$, a lower bound on $\mathbb{E}_{\mathbb{P}} f(\boldsymbol{x},\boldsymbol{z})$ is given by $g_L(\boldsymbol{x})$ ($g_U(\boldsymbol{x})$ respectively). The upper and lower bound give a closed interval in which Val lies. That is, for the convex case it is guaranteed that Val lies in the interval $[g_L(\boldsymbol{x}), g_U(\boldsymbol{x})]$ and in the concave case in the interval $[g_U(\boldsymbol{x}), g_L(\boldsymbol{x})]$. The above result applies also to the case of the ambiguous constraints (4.2).

There are two difficulties associated with the bounds (4.20) and (4.21). One is the computational difficulty: when $n_{\boldsymbol{z}}$ is large, formulas (4.20) and (4.21) include an exponential number of terms. Second is the independence assumption on z_i's:

when the independence hypothesis is rejected, the solutions obtained using (4.20) and (4.21) might underperform significantly. In Section 4.4 we discuss a wide class of functions $f(\boldsymbol{x}, \boldsymbol{z})$ for which both of these difficulties are alleviated. Here, we discuss cases where these difficulties are not present or can be alleviated using existing techniques.

Dimensionality.

- In certain applications the number $n_{\boldsymbol{z}}$ of random variables is small (less than 10).
- An important special case is when $f(\boldsymbol{x}, \boldsymbol{z})$ is a sum of functions

$$f(\boldsymbol{x}, \boldsymbol{z}) = \sum_{j=1}^{n_c} f^{(j)}\left(\boldsymbol{x}, \boldsymbol{z}^{(j)}\right),$$

  where $f^{(j)}(\boldsymbol{x}, \cdot)$ have small numbers n_j of uncertain variables.
- An important special case is the function $f(\boldsymbol{x}, \boldsymbol{z}) = \exp(\boldsymbol{x}^T \boldsymbol{z})$. Upper bounds on moment generating functions $\mathbb{E}\exp(\boldsymbol{x}^T \boldsymbol{z})$ are a key tool in constructing safe tractable approximations of chance constraints. As we show in Section 4.5, the properties of the $\exp(\cdot)$ allow for a simple, closed-form formula for its worst-case expectation under (μ, d) information and for which the number of terms is linear in $n_{\boldsymbol{z}}$.
- If the dimensionality remains an issue, for problems with linear and piecewise linear $f(\boldsymbol{x}, \cdot)$ one can use, for example, the Stochastic Decomposition method (Higle and Sen 1996) where scenarios (in our case support points) are added iteratively until the current model is a good enough approximation of the original model. In cutting-plane methods, the verification of the ambiguous constraint can exploit the tree-structure of the worst-case distribution support. In this tree structure, each outcome of z_1 leads to 3 (or 2 for the concave case) outcomes of z_2, each of these leads to another 3 outcomes of z_3 etc. Then, one can determine if the constraint holds already after investigating the first few layers of the tree, which may lead to a verification of much less than all $3^{n_{\boldsymbol{z}}}$ scenarios. Other approximate approaches are the sample average approximation (Shapiro et al. 2009) or the scenario reduction technique (Dupačová et al. 2003).

Dependence.

If the random uncertain vector $\boldsymbol{z}$ contains dependent components, it can be decomposed by means of factor analysis, for example, based on Principal Component Analysis (see Jolliffe (2002)), into linear combinations of a limited number of uncorrelated factors. For example, in a situation of portfolio optimization problem with

25 assets, it is natural to decompose them into 3-4 uncorrelated risk factors (see, for example Baillie et al. (2002)), whose empirical distribution provides information also about their support, means, and MADs. Even though uncorrelatedness can be much weaker than independence, such a technique is often a practical solution.

4.3.2 The use of the Ben-Tal and Hochman (1972) bounds in some general applications

In this section we present three cases where the reformulations of the worst-case expected feasibility constraints presented in Section 4.3.1 can be used.

Average-case enhancement of RO solutions. The first application lies in finding worst-case-optimal solutions with good average-case performance to the following RO problem:

$$\begin{aligned} \min_{\boldsymbol{x},t} \quad & t \\ \text{s.t.} \quad & \sup_{z\in\mathcal{Z}} f(\boldsymbol{x},\boldsymbol{z}) \leq t, \\ & \sup_{z\in\mathcal{Z}} g_i(\boldsymbol{x},\boldsymbol{z}) \leq 0, \quad i=1,\ldots,m. \end{aligned} \tag{4.22}$$

It happens frequently that there exist multiple optimal solutions to (4.22), see Iancu and Trichakis (2013), De Ruiter et al. (2016). Whereas the worst-case performance of such solutions is the same, their average-case performance may differ dramatically. For that reason, once the optimal value $\bar{t}$ for (4.22) is known, a second optimization step may be used to select one of the optimal solutions to provide good average-case behavior. Since the results of Ben-Tal and Hochman (1972) provide exact bounds on the worst-case expectations, they can be used in such a step. In the following, we describe such a two-step procedure:

1. Solve problem (4.22) and denote its optimal value by $\bar{t}$.
2. Solve the following problem, minimizing the worst-case expectation of the objective value, with the worst-case value of $f(\boldsymbol{x},\boldsymbol{z})$ less than or equal to $\bar{t}$:

$$\begin{aligned} \min_{\boldsymbol{x},u} \quad & u \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{x},\boldsymbol{z}) \leq u \\ & \sup_{z\in\mathcal{Z}} f(\boldsymbol{x},\boldsymbol{z}) \leq \bar{t}, \\ & \sup_{z\in\mathcal{Z}} g_i(\boldsymbol{x},\boldsymbol{z}) \leq 0, \quad i=1,\ldots,m. \end{aligned} \tag{4.23}$$

In case of multiple optimal solutions to (4.22), the two-step procedure is expected to select the optimal solution with good average-case performance for its focus on

the worst-case expectation among the best worst-case solutions. If the uncertainty is present only in the constraints involving functions $g_i(\cdot,\cdot)$, a similar two-step approach can be designed to maximize the worst-case expected slack in the worst-case constraints in (4.22), see Iancu and Trichakis (2013). We note that following the theory of Iancu and Trichakis (2013), there might exist multiple optimal solutions to (4.23) and one may need to include another 'enhancement step' to choose among them.

An alternative approach to enhancing robust solutions is to sample a number S of scenarios for $\boldsymbol{z}$ to find a solution that optimizes the average of the objective value over the sample.[1] This approach, however, has as shortcoming that the outcome might depend on the choice of sample size S and the sample itself. For that reason, the DRO methods can provide a good alternative to enhancing the quality of RO solutions. In our paper, we test the application of the (μ, d) bounds to enhance average-case performance in an inventory management problem in Section 4.3.4.

Implementation error. The second application we consider is when the decision variables cannot be implemented with the designed value due to implementation error in the following problem:

$$\begin{aligned} \min_{\boldsymbol{x},t} \quad & t \\ \text{s.t.} \quad & f(\boldsymbol{x}) \leq t, \\ & g_i(\boldsymbol{x}) \leq 0, \quad i = 1,\ldots,m. \end{aligned} \tag{4.24}$$

In case of the existence of an additive implementation error $\boldsymbol{z}$ the implemented value is $\boldsymbol{x} = \overline{\boldsymbol{x}} + \boldsymbol{z}$, where $\overline{\boldsymbol{x}}$ is the designed value and $\boldsymbol{z} = (z_1,\ldots,z_{n_{\boldsymbol{x}}})^T$ is the error. Then, the problem becomes:

$$\begin{aligned} \min_{\overline{\boldsymbol{x}},t} \quad & t \\ \text{s.t.} \quad & \sup_{z\in\mathcal{Z}} f(\overline{\boldsymbol{x}} + \boldsymbol{z}) \leq t, \\ & \sup_{z\in\mathcal{Z}} g_i(\overline{\boldsymbol{x}} + \boldsymbol{z}) \leq 0, \quad i = 1,\ldots,m. \end{aligned} \tag{4.25}$$

Since $f(\boldsymbol{x})$ is convex in $\boldsymbol{x}$, in (4.25) the function $f(\overline{\boldsymbol{x}}+\boldsymbol{z})$ is convex in $\boldsymbol{z}$. For that reason, optimization of the worst-case value of the objective function could be difficult, as typically RO techniques rely on the constraint being concave in the uncertain parameter (see Ben-Tal et al. (2009, 2015)). Therefore, optimizing the worst-case values of convex constraints under implementation error is a problem leading to computational intractability, apart from special cases such as linear constraints (see Ben-Tal et al.

[1] As a special case, one can choose only one scenario, corresponding to the nominal values of the uncertain parameters (Iancu and Trichakis 2013).

(2015)) or (conic) quadratic constraints with simultaneously diagonizable quadratic forms defining the constraint and the uncertainty set for the error (see Ben-Tal and den Hertog (2011)).

Because of the above, it may be an alternative to optimize the worst-case expectation of the objective function, for which our DRO method applies under the corresponding distributional assumptions on $\boldsymbol{z}$, i.e., that the ambiguity set for the distribution of $\boldsymbol{z}$ is $\mathcal{P}_{(\mu,d)}$. Then, the problem becomes:

$$\begin{aligned} \min_{\overline{\boldsymbol{x}},t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\overline{\boldsymbol{x}}+\boldsymbol{z}) \leq t, \\ & \sup_{z\in\mathcal{Z}} g_i(\overline{\boldsymbol{x}}+\boldsymbol{z}) \leq 0, \quad i=1,\ldots,m. \end{aligned} \tag{4.26}$$

The first constraint in (4.26) is convex in $\boldsymbol{z}$ and one can apply the reformulation (4.20). For (4.26) to be tractable, the functions $g_i(\overline{\boldsymbol{x}}+\boldsymbol{z})$ need to be affine in $\boldsymbol{z}$ or belong to one of the special cases considered in Ben-Tal and den Hertog (2011). Similarly, one can reformulate a problem where multiplicative error occurs, i.e., where $\boldsymbol{x} = (\overline{x}_1 z_1, \ldots, \overline{x}_{n_{\boldsymbol{x}}} z_{n_{\boldsymbol{x}}})^T$.

Convex constraints and linear decision rules. The third application of our DRO approach comes when the constraints of a problem are convex in z as a result of applying linear decision rules. To show how such a situation occurs, we consider a two-stage RO problem:

$$\begin{aligned} \min_{\boldsymbol{x}_1,\boldsymbol{x}_2,t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{x}_1,\boldsymbol{x}_2(\boldsymbol{z}),\boldsymbol{z}) \leq t \\ & \sup_{z\in\mathcal{Z}} g_i(\boldsymbol{x}_1,\boldsymbol{x}_2(\boldsymbol{z}),\boldsymbol{z}) \leq 0, \quad i=1,\ldots,m, \end{aligned} \tag{4.27}$$

where $\boldsymbol{x}_1 \in \mathbb{R}^{n_{\boldsymbol{x}_1}}$ is implemented at before $\boldsymbol{z}$ is known (time 1) and $\boldsymbol{x}_2 \in \mathbb{R}^{n_{\boldsymbol{x}_2}}$ is implemented after $\boldsymbol{z}$ is known (time 2), i.e. $\boldsymbol{x}_2 = \boldsymbol{x}_2(z)$. In such cases, it is possible to define the time-2 decisions as a linear function $\boldsymbol{x}_2(z) = \boldsymbol{v} + \boldsymbol{V}\boldsymbol{z}$ of the uncertain parameter z (see Ben-Tal et al. (2004)), to provide adjustability of decisions at time

2.[2] The problem is then:

$$\begin{aligned} \min_{\boldsymbol{x}_1,\boldsymbol{v},\boldsymbol{V},t} \quad & t \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{x}_1, \boldsymbol{v} + \boldsymbol{V}\boldsymbol{z}, \boldsymbol{z}) \leq t \\ & \sup_{\boldsymbol{z}\in\mathcal{Z}} g_i(\boldsymbol{x}_1, \boldsymbol{v} + \boldsymbol{V}\boldsymbol{z}, \boldsymbol{z}) \leq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{4.28}$$

Since $f(\boldsymbol{x}_1, \boldsymbol{x}_2(\boldsymbol{z}), \boldsymbol{z})$ is convex in $\boldsymbol{x}_2$, the first constraint in (4.28) may also be convex in $\boldsymbol{z}$. In such a case, a further reformulation of problem (4.27) can be conducted as in Section 4.3.2. We combine linear decision rules with (μ, d) information in the inventory problem of Sections 4.3.3 and 4.3.4.

4.3.3 Application 1: Inventory management - average case performance

Introduction. In this section we consider an application of the (μ, d) method to minimization of the average-case costs in inventory management. The main research questions are:

1. How does minimizing the worst-case expectation affect the best-case expectation under the given distributional assumptions?

2. What is the average-case performance of solutions minimizing the worst-case expectation compared to the robust solutions, minimizing the worst-case outcome of the objective values?

To answer them, we adapt the numerical example from Ben-Tal et al. (2005) with a single product and where inventory is managed periodically over T periods. At the beginning of each period t the decision maker has an inventory of size x_t and he orders a quantity q_t for unit price c_t. The customers then place their demands z_t. The retailer's status at the beginning of the planning horizon is given through the parameter x_1 (initial inventory). Apart from the ordering costs, the following costs are incurred over the planning horizon:

- holding cost $h_t \max\{0, x_{t+1}\}$, where h_t are the unit holding costs,

- shortage costs $p_t \max\{0, x_{t+1}\}$, where p_t are the unit shortage costs.

[2] One may also use other decision rules. However, we limit ourselves to the analysis of the linear case as the linear decision rules are very often a powerful enough tool, see Bertsimas et al. (2011). Moreover, the (non)convexity of the problem resulting from application of linear decision rules is easy to verify, see Boyd and Vandenberghe (2004).

Inventory x_{T+1} left at the end of period T has a unit salvage value s. Also, one must impose $h_T - s \geq -p_T$ to maintain the problem's convexity. Practical interpretation of this constraint is that in the last period it is more profitable to satisfy the customer demand rather than to be left with excessive amount of inventory. The constraints in the model include:

- balance equations linking the inventory in each period to the inventory, order quantity, and demand in the preceding period,
- upper and lower bounds on the order quantities in each period $L_t \leq q_t \leq U_t$,
- upper and lower bounds on cumulative order quantities in each period $\widehat{L}_t \leq \sum_{\tau=1}^{t} q_\tau \leq \widehat{U}_t$.

With ordering decisions $\boldsymbol{q}(\boldsymbol{z}) = (q_1, q_2(\boldsymbol{z}^1), \ldots, q_T(\boldsymbol{z}^{T-1}))^T$, where $\boldsymbol{z}^t = (z_1, \ldots, z_t)^T$, the objective function value for a given demand vector $\boldsymbol{z}$ is

$$\begin{aligned} f(\boldsymbol{q}(\boldsymbol{z}), \boldsymbol{z}) = \quad & \sum_{t=1}^{T} \left(c_t q_t(\boldsymbol{z}^{t-1}) + h_t \max\left\{ x_{t+1}(\boldsymbol{z}^t), 0 \right\} + p_t \max\left\{ -x_{t+1}(\boldsymbol{z}^t), 0 \right\} \right) \\ & - s \max\left\{ x_{T+1,0}(\boldsymbol{z}^T) \right\}. \end{aligned}$$

The optimization problem to be solved is given by the following, two-variant formulation where the minimized quantity is the worst-case value or the worst-case expectation of the objective function:

$$\begin{aligned} \min_{\boldsymbol{q}(\cdot), u} \quad & u \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \text{ or } \sup_{\boldsymbol{z} \in \mathcal{Z}} f(\boldsymbol{q}(\boldsymbol{z}), \boldsymbol{z}) \leq u \\ & x_{t+1}(\boldsymbol{z}^t) = x_t(\boldsymbol{z}^{t-1}) + q_t - z_t, \quad t = 1, \ldots, T \\ & L_t \leq q_t(\boldsymbol{z}^{t-1}) \leq U_t, \quad t = 1, \ldots, T \\ & \widehat{L}_t \leq \sum_{\tau=1}^{t} q_\tau(\boldsymbol{z}^{\tau-1}) \leq \widehat{U}_t, \quad t = 1, \ldots, T, \end{aligned} \tag{4.29}$$

where $\mathcal{Z}$ is the uncertainty set for $\boldsymbol{z}$ and $\mathcal{P}$ is the ambiguity set of probability distributions with support being a subset of $\mathcal{Z}$. The objective function in (4.29) has the sum-of-maxima form which typically is problematic in RO due to the difficulty of maximizing a convex function. It is of no concern as the Ben-Tal and Hochman (1972) results only require that the function at hand is convex in the uncertain parameter.

We assume that the uncertainty set $\mathcal{Z}$ is $\mathcal{Z} = \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_T$, where $\mathcal{Z}_t = [a_t, b_t]$, $t = 1, \ldots, T$, which corresponds to $\boldsymbol{z}$ being a random variable with independent components. The worst-case form of problem (4.29) has to be solved by enumerating all vertices of the uncertainty set $\mathcal{Z}$. For the worst-case expectation form of (4.29) we

Table 4.1 – Ranges for parameter sampling in the inventory experiment.

Parameter	Range	Parameter	Range
a_t	$[0, 20]$	x_1	$[20, 50]$
b_t	$[a_t, a_t + 100]$	L_t	0
c_t, p_t	$[0, 10]$	U_t	$[50, 70]$
h_t	$[0, 5]$	$\hat{L}_t$	0
s	0	$\hat{U}_t$	$0.8 \sum_{t=1}^{T} U_t$

assume that $\mu_t = \frac{a_t+b_t}{2}$, and that $d_t = \mathbb{E}_{\mathbb{P}}|z_t - \mu_t| = \theta(b_t - a_t)$, yielding the following ambiguity set:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P} : \quad \text{supp}(\mathbb{P}) \subset [\boldsymbol{a}, \boldsymbol{b}], \quad \mathbb{E}_{\mathbb{P}}\boldsymbol{z} = \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}}\,|\boldsymbol{z} - \boldsymbol{\mu}| = \boldsymbol{d}, \quad z_i \perp\!\!\!\perp z_j \quad \forall i \neq j\}\,,$$

where $\boldsymbol{a} = (a_1, \ldots, a_T)^T$, $\boldsymbol{b} = (b_1, \ldots, b_T)^T$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_T)^T$, $\boldsymbol{d} = (d_1, \ldots, d_T)^T$. The ordering decisions are assumed to be linear functions of the past demand: $q_{t+1}(\boldsymbol{z}^t) = q_{t+1,0} + \sum_{j=1}^{t} q_{t+1,j} z_j$ and require that $q_{t+1}(\boldsymbol{z}^t) \geq 0$ for all $\boldsymbol{z} \in \mathcal{Z}$, for $t = 2, \ldots, T + 1$. We solve the following two variants of problem (4.29):

- RO solution - the objective function in (4.29) is preceded by $\sup_{\boldsymbol{z}\in\mathcal{Z}}$,
- (μ, d) solution - the objective function in (4.29) is preceded by $\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}$.

We conduct an experiment with $T = 6$ and 50 problem instances. We set $\theta = 0.25$, corresponding to the mean absolute deviation of the uniform distribution. The ranges for the uniform sampling of parameters are given in Table 4.1.

Upper and lower bounds for the expectation of the objective function. We consider now the first research question of this section. For each inventory problem instance and the optimal solution $\overline{\boldsymbol{q}}(\cdot)$, we compute the following quantities:

- the worst-case expected cost under (μ, d) information: $\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\overline{\boldsymbol{q}}(\boldsymbol{z}), \boldsymbol{z})$
- the best-case expected cost $\inf_{\mathbb{P}\in\mathcal{P}_{(\mu,d,\beta)}} \mathbb{E}_{\mathbb{P}} f(\overline{\boldsymbol{q}}(\boldsymbol{z}), \boldsymbol{z})$ with three possibilities for the skewness of the demand distribution, i.e., with $\beta_t = \beta \in \{0.25, 0.5, 0.75\}$, corresponding to left-skewness, symmetry, and right-skewness of the demand distribution in all periods, respectively.

The two values provide us with information about the interval within which the expected objective function value lies under three different assumptions on the parameter β. Additionally, for each solution we compute the worst-case cost $\sup_{\boldsymbol{z}\in\mathcal{Z}} f(\overline{\boldsymbol{q}}(\boldsymbol{z}), \boldsymbol{z})$ to verify how the minimization of the worst-case expectation affects the worst-case performance of the solution.

Table 4.2 – Results of the inventory management - worst-case costs and ranges for the expectation of the objective over $\mathcal{P}_{(\mu,d,\beta)}$. All numbers are averages.

Objective type	β	Minimum cost	
		RO	(μ, d)
Worst-case value	-	1950	2384
Expectation range	0.25	[1255,1280]	[1004,1049]
Expectation range	0.5	[1223,1280]	[970,1049]
Expectation range	0.75	[1230,1280]	[994,1049]

Table 4.2 presents the results. As can be expected, the RO solution yields the best worst-case value of 1950 which is far better than the (μ, d) solution, whose worst-case value is 2384. Rows 2 to 4 show that the (μ, d) solution not only yields better upper bounds on the expected value of the solution, but also leads to an improvement of the best-case expectation for all β. For example, for $\beta = 0.5$ the interval for the expected cost related to the RO solution is given by $[1255, 1280]$, whereas for the (μ, d) solution it is $[970, 1049]$. That means that the worst-case expectation obtained by the (μ, d) solution is better than the worst-case expectation obtained by the RO solution.

Simulation results. We now answer the second research question by conducting a simulation study. Since the solutions are obtained with different objective functions, comparing their average-case performance in a 'fair' way is difficult. We compare their performance using two samples of demand vectors $\hat{\boldsymbol{z}}$:

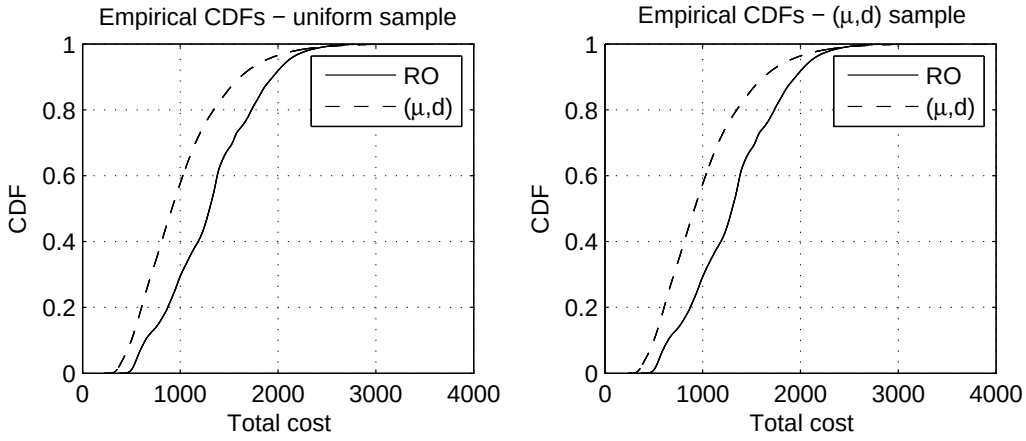
- uniform sample - demand scenarios $\hat{\boldsymbol{z}}$ are sampled from a uniform distribution on $\mathcal{Z}$,
- (μ, d) sample - demand scenarios $\hat{\boldsymbol{z}}$ are sampled from a distribution $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$. That is, first, a discretized distribution $\hat{\mathbb{P}} \in \mathcal{P}_{(\mu,d)}$ is sampled using the hit-and-run method. This method is implemented here as follows. For the $[0, 1]$-interval we construct a grid of 50 equidistant points. For a fixed (μ, d) the set of probability masses assigned to these points satisfying the μ and d values is a polytope. We sample 10 probability distributions uniformly from this polytope with the classical hit-and-run method (mixing algorithm) of Smith (1984), where we choose the starting point to be the analytic center of the polytope and we use only every 20th vector sampled with the mixing algorithm. Then, each component of the vector $\hat{\boldsymbol{z}}$ is sampled randomly from a randomly chosen distribution $\hat{\mathbb{P}}$.

For each instance, we sample 10^4 demand scenarios, with both of the sampling methods. Table 4.3 presents the results. The averages of the objective function values over the two sample types over all instance are put in bold. For example (row 1),

Table 4.3 – Simulation results for the first inventory problem. Numbers in brackets denote the % change compared to the RO solution.

Objective type	Demand sample type	Cost	
		RO	(μ, d)
Objective mean	Uniform sample	**1230**	**994 (-19.6%)**
Objective standard deviation	Uniform sample	157	259 (+65%)
Objective mean	(μ, d) sample	**1246**	**1003 (-19.5%)**
Objective standard deviation	(μ, d) sample	160	265 (+65.6%)

Figure 4.1 – Empirical cumulative distribution functions of the total costs of the solutions under the uniform sample (left) and (μ, d) sample (right). Aggregated from all problem instances.



the (μ, d) solutions perform better on average in the uniform sample, with values 994 and 1230, respectively. A similar observation holds for the (μ, d) sample (row 3). In Figure 4.1 we present a comparison of the empirical cumulative distribution functions of the total costs incurred by the RO and (μ, d) solutions. In both samples we can see that the costs of the RO solution stochastically dominate over the ones from the (μ, d) solutions. Thus, we conclude that the (μ, d) solutions are superior to the RO solution.

4.3.4 Application 2: Inventory management - enhancement of RO solutions

With the good average-case performance of the (μ, d) solutions in the previous experiment, we investigate now the following question: can the (μ, d) method be used to enhance the average-case performance of RO solutions? That is, is it possible, in cases where there are multiple optimal solutions to the RO problem, to find the worst-case optimal solution that has a better average cost than the initial worst-case optimal solution? To verify this, for each of the problem instances of the previous subsection we apply the two-step procedure of Section 4.3.2.

Table 4.4 – Results of the inventory management - enhancement of RO solutions example. All numbers are averages. Numbers in brackets denote the % change compared to the initial solution with no enhancement (first column).

Objective type		Cost			
	Enhancement type	-	(μ, d)	Sample	Nominal
Objective mean	Uniform sample	**1230**	**1168 (-5.04%)**	1168 (-5.04%)	1180 (-4.06%)
Objective standard deviation	Uniform sample	157	158 (+0.63%)	156 (-0.63%)	161 (+2.54%)
Objective mean	(μ, d) sample	**1246**	**1172 (-5.93%)**	1172 (-5.93%)	1184 (-4.97%)
Objective standard deviation	(μ, d) sample	160	161 (+0.62%)	160 (0.00%)	164 (+2.50%)

We consider three enhancement types, corresponding to three different objective functions:

- (μ, d) enhancement: $\min \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} f(\boldsymbol{q}(\boldsymbol{z}), \boldsymbol{z})$,
- sample enhancement: $\min \frac{1}{S} \sum_{j=1}^{S} f(\boldsymbol{q}(\hat{\boldsymbol{z}}_j), \hat{\boldsymbol{z}}_j)$, where $\hat{\boldsymbol{z}}_j$ are $S = 200$ demand scenarios sampled uniformly from $\mathcal{Z}$,
- nominal enhancement: $\min f(\boldsymbol{q}(\boldsymbol{\mu}), \boldsymbol{\mu})$ considered by Iancu and Trichakis (2013).

Table 4.4 presents the results. In the uniform sample (row 1) the (μ, d)-enhanced solution yields an average cost of 1168, compared to 1230 for the non-enhanced solution, that is 5.04% less. For the (μ, d) sample (row 3) the corresponding number is 5.93%. The nominal enhancement turns out to be slightly worse than the (μ, d) and sample enhancements, compare e.g. the means 1180 and 1168 for the uniform sample and the higher standard deviations of the nominal enhancement.

In Figure 4.2 we present the empirical cumulative distribution functions of the non-enhanced and (μ, d)-enhanced solutions in a sample problem in the uniform demand sample. In this plot, it is clear that the total cost incurred by the non-enhanced solutions stochastically dominates the one from the (μ, d) enhanced solutions.

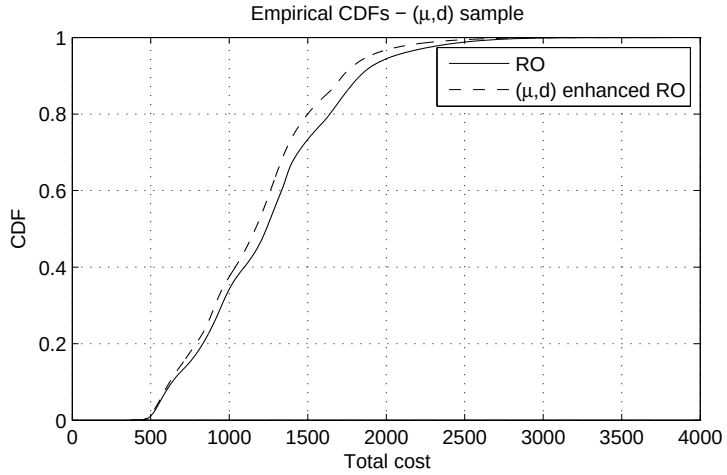
4.4 Extension - aggregated random vectors

4.4.1 Introduction

Up to now, we have been deriving exact worst-case expectations which (i) relied on the assumption of independence of the components of $\boldsymbol{z}$, and (ii) resulted in bounds involving $3^{n_{\boldsymbol{z}}}$ terms. In this section, we consider practically relevant cases where both of these difficulties can be alleviated.

In many cases, uncertainty appears in a linearly aggregated way such as $y = \boldsymbol{a}^T \boldsymbol{z}$ or $y(\boldsymbol{x}) = \boldsymbol{a}(\boldsymbol{x})^T \boldsymbol{z}$. Then, instead of considering the worst-case expectation of $f(\boldsymbol{x}, \boldsymbol{a}^T \boldsymbol{z})$

Figure 4.2 – Empirical cumulative distribution functions of the total costs in simulation in the uniform demand sample for the non-enhanced RO solution and the (μ, d)-enhanced RO solution. Aggregated from all problem instances.



or $f(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z})$ (or, more generally, $\sum_i f(\boldsymbol{x}, \boldsymbol{a}_i^T\boldsymbol{z})$ or $\sum_i f(\boldsymbol{a}_i(\boldsymbol{x})^T\boldsymbol{z})$), it is possible to consider the worst-case expectations of $f(\boldsymbol{x}, y)$, where y is a single-dimensional ambiguous random variable equal either to $\boldsymbol{a}^T\boldsymbol{z}$ or to $\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z}$ while the uncertainty is still specified in terms of the entire vector $\boldsymbol{z}$. Without loss of generality, we assume that the support of $\boldsymbol{z}$ is given by $\|\boldsymbol{z}\|_\infty \leq 1$.

4.4.2 Fixed vector $\boldsymbol{a}$

We begin our analysis with the case where the vector $\boldsymbol{a}$ is not dependent on the decision variables, motivated by the following example.

Example 4.1 *In the inventory problem of Section 4.3.3 the holding and backlogging costs at time $t+1$ depend on the state of inventory x_{t+1}. If the ordering decisions $q_t(\boldsymbol{z}^{t-1})$ are static (non-adjustable), then:*

$$x_{t+1} = x_1 + \sum_{s=1}^{t} q_s - \sum_{s=1}^{t} z_s = x_1 + \sum_{s=1}^{t} q_s - \mathbf{1}^T\boldsymbol{z}^t,$$

and the aggregated random variable is $y = \mathbf{1}^T\boldsymbol{z}^t$ not depending on the decision variables q_t. ■

In order to use the results of Ben-Tal and Hochman (1972) to construct the worst-case expectation of $f(\boldsymbol{x}, y)$, we need to extract the distributional information on $y = \boldsymbol{a}^T\boldsymbol{z}$ from the information on $\boldsymbol{z}$. Then, we have that

$$\begin{aligned} \operatorname{supp}(\boldsymbol{a}^T\boldsymbol{z}) = [\min_{\boldsymbol{z}} \boldsymbol{a}^T\boldsymbol{z}, \max_{\boldsymbol{z}} \boldsymbol{a}^T\boldsymbol{z}] &= [-\|\boldsymbol{a}\|_1, \|\boldsymbol{a}\|_1] \\ \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{a}^T\boldsymbol{z}\right) &= \boldsymbol{a}^T\boldsymbol{\mu}. \end{aligned} \tag{4.30}$$

As for the MAD $M(\boldsymbol{a}^T\boldsymbol{z})$, only upper bounds on $M(\boldsymbol{a}^T\boldsymbol{z})$ can be obtained in terms of information on $\boldsymbol{z}$. Nevertheless, any upper bound on $M(\boldsymbol{a}^T\boldsymbol{z})$ will generate an upper bound on $\mathbb{E}_{\mathbb{P}}f(\boldsymbol{x}, y)$ due to the fact that the Ben-Tal and Hochman (1972) upper bound (4.8) is a nondecreasing function of d, as stated by the following Proposition.

Proposition 4.1 *The worst-case expectation (4.8) is a nondecreasing function of d.*

Proof. The worst-case expectation (4.8) is:

$$F(d) = \frac{d}{2(\mu - a)}f(a) + \left(1 - \frac{d}{2(\mu - a)} - \frac{d}{2(b - \mu)}\right) f(\mu) + \frac{d}{2(b - \mu)}f(b).$$

We have

$$F'(d) = \frac{1}{2(\mu - a)}f(a) - \left(\frac{1}{2(\mu - a)} + \frac{1}{2(b - \mu)}\right) f(\mu) + \frac{1}{2(b - \mu)}f(b) \geq 0.$$

Multiplying the last inequality by $2(b - \mu)(\mu - a)/(b - a)$ and using

$$\mu = \frac{b - \mu}{b - a}a + \frac{\mu - a}{b - a}b$$

we obtain the inequality:

$$\frac{b - \mu}{b - a}f(a) + \frac{\mu - a}{b - a}f(b) \geq f\left(\frac{\mu - a}{b - a}b + \frac{b - \mu}{b - a}a\right),$$

which is valid by convexity of $f(\cdot)$. ■

In the following, we present four ways to obtain upper bounds on the MAD $M(\boldsymbol{a}^T\boldsymbol{z})$. The first three of them do not use the assumption of independence of the components of $\boldsymbol{z}$ and are hence computable in polynomial time. The last one, on the other hand, requires the independence of z_i's, but the computation involving $3^{n_{\boldsymbol{z}}}$ terms can be done in a pre-processing step, without affecting the size of the optimization problem.

No independence - simple bounds. Two upper bounds that we use here are:

$$\mathbb{E}_{\mathbb{P}}\left|y - \boldsymbol{a}^T\boldsymbol{\mu}\right| = \mathbb{E}_{\mathbb{P}}|\sum_{i=1}^{n_{\boldsymbol{z}}} a_i z_i - \sum_{i=1}^{n_{\boldsymbol{z}}} a_i\mu_i| \leq \sum_{i=1}^{n_{\boldsymbol{z}}} \mathbb{E}_{\mathbb{P}}\,|a_i z_i - a_i\mu_i| = \sum |a_i|d_i = |\boldsymbol{a}|^T\boldsymbol{d}, \tag{4.31}$$

which gives the following ambiguity set for the distribution of y:

$$\mathcal{P}_y^{\boldsymbol{d}} = \left\{\mathbb{P}_y : \quad \text{supp}(\mathbb{P}_y) \subseteq [-\|\boldsymbol{a}\|_1, \|\boldsymbol{a}\|_1]], \quad \mathbb{E}_{\mathbb{P}_y}y = \boldsymbol{a}^T\boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}_y}|y - \boldsymbol{a}^T\boldsymbol{\mu}| \leq |\boldsymbol{a}|^T\boldsymbol{d}\right\}, \tag{4.32}$$

and the second bound, based on the covariance matrix $\Sigma_{\boldsymbol{z}}$, is

$$\mathbb{E}_{\mathbb{P}} \left| y - \boldsymbol{a}^T \boldsymbol{\mu} \right| \leq \sqrt{\mathbb{E}\left(y - \boldsymbol{a}^T \boldsymbol{\mu}\right)^2} = \sqrt{\mathbb{E}\left(\boldsymbol{a}^T \boldsymbol{z} - \boldsymbol{a}^T \boldsymbol{\mu}\right)^2} = \sqrt{\mathrm{Var}(\boldsymbol{a}^T \boldsymbol{z})} = \sqrt{\boldsymbol{a}^T \Sigma_{\boldsymbol{z}} \boldsymbol{a}}, \tag{4.33}$$

which gives the following ambiguity set for the distribution of y:

$$\mathcal{P}_y^{\mathrm{Cov}} = \left\{ \mathbb{P}_y : \quad \mathrm{supp}(\mathbb{P}_y) \subseteq [-\|\boldsymbol{a}\|_1, \|\boldsymbol{a}\|_1|, \quad \mathbb{E}_{\mathbb{P}_y} y = \boldsymbol{a}^T \boldsymbol{\mu}, \quad \mathbb{E}_{\mathbb{P}_y} |y - \boldsymbol{a}^T \boldsymbol{\mu}| \leq \sqrt{\boldsymbol{a}^T \Sigma_{\boldsymbol{z}} \boldsymbol{a}} \right\}. \tag{4.34}$$

Ben-Tal and Hochman (1972) bounds obtained using ambiguity sets (4.32) and (4.34) do not require the components of $\boldsymbol{z}$ to be independent and require only three terms, thus using (4.32) :

$$\begin{aligned} \sup_{\mathbb{P}_y \in \mathcal{P}_y^{\boldsymbol{d}}} \mathbb{E}_{\mathbb{P}_q} f(\boldsymbol{x}, y) \quad \leq \quad & \frac{|\boldsymbol{a}|^T \boldsymbol{d}}{2(\boldsymbol{a}^T \boldsymbol{\mu} + \|\boldsymbol{a}\|_1)} f(\boldsymbol{x}, -\|\boldsymbol{a}\|_1) + \\ & + \left(1 - \frac{|\boldsymbol{a}|^T \boldsymbol{d}}{2(\boldsymbol{a}^T \boldsymbol{\mu} + \|\boldsymbol{a}\|_1)} - \frac{|\boldsymbol{a}|^T \boldsymbol{d}}{2(\|\boldsymbol{a}\|_1 - \boldsymbol{a}^T \boldsymbol{\mu})} \right) f(\boldsymbol{x}, \boldsymbol{a}^T \boldsymbol{\mu}) + \\ & + \frac{|\boldsymbol{a}|^T \boldsymbol{d}}{2(\|\boldsymbol{a}\|_1 - \boldsymbol{a}^T \boldsymbol{\mu})} f(\boldsymbol{x}, \|\boldsymbol{a}\|_1). \end{aligned}$$

Since $\boldsymbol{a}$ does not depend on $\boldsymbol{x}$, the resulting expression is convex in $\boldsymbol{x}$.

No independence - exact bound using Wiesemann et al. (2014). It is also possible to obtain an exact upper bound on the MAD of y using the results of Wiesemann et al. (2014), which requires, though, solving an optimization problem. We present it on the example of (μ, d) information about $\boldsymbol{z}$ but it can also use the mean-covariance and some other types of information on $\boldsymbol{z}$. The problem to solve is:

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left| \boldsymbol{a}^T \boldsymbol{z} - \boldsymbol{a}^T \boldsymbol{\mu} \right|. \tag{4.35}$$

It can be proved (see Appendix 4.F to this paper) using Theorem 1 of Wiesemann et al. (2014) that (4.35) is equivalent to:

$$\begin{aligned}
\min_{\boldsymbol{\phi}_1,\boldsymbol{\phi}_2\geq 0,w,\boldsymbol{\beta},\kappa} \quad & w \\
\text{s.t.} \quad & \boldsymbol{b}^T\boldsymbol{\beta}+\kappa \leq w \\
& \boldsymbol{c}^T\boldsymbol{\phi}_1 - \boldsymbol{a}^T\boldsymbol{\mu} \leq \kappa \\
& \boldsymbol{c}^T\boldsymbol{\phi}_2 + \boldsymbol{a}^T\boldsymbol{\mu} \leq \kappa \\
& \boldsymbol{C}^T\boldsymbol{\phi}_1 + \boldsymbol{A}^T\boldsymbol{\beta} = \boldsymbol{a} \\
& \boldsymbol{C}^T\boldsymbol{\phi}_2 + \boldsymbol{A}^T\boldsymbol{\beta} = -\boldsymbol{a} \\
& \boldsymbol{D}^T\boldsymbol{\phi}_1 + \boldsymbol{B}^T\boldsymbol{\beta} = 0 \\
& \boldsymbol{D}^T\boldsymbol{\phi}_2 + \boldsymbol{B}^T\boldsymbol{\beta} = 0,
\end{aligned} \tag{4.36}$$

where $\boldsymbol{A},\boldsymbol{B}\in\mathbb{R}^{2n_{\boldsymbol{z}}\times n_{\boldsymbol{z}}}$, $\boldsymbol{b}\in\mathbb{R}^{2n_{\boldsymbol{z}}}$, $\boldsymbol{C},\boldsymbol{D}\in\mathbb{R}^{6n_{\boldsymbol{z}}\times n_{\boldsymbol{z}}}$, and $\boldsymbol{c}\in\mathbb{R}^{6n_{\boldsymbol{z}}}$ are defined as:

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{B}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{I}\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}\boldsymbol{\mu}\\ \boldsymbol{d}\end{bmatrix},\quad \boldsymbol{C}=\begin{bmatrix}\boldsymbol{I}\\ -\boldsymbol{I}\\ \boldsymbol{I}\\ -\boldsymbol{I}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix},\quad \boldsymbol{D}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ -\boldsymbol{I}\\ -\boldsymbol{I}\\ \boldsymbol{I}\\ -\boldsymbol{I}\end{bmatrix},\quad \boldsymbol{c}=\begin{bmatrix}\boldsymbol{1}\\ \boldsymbol{1}\\ \boldsymbol{\mu}\\ -\boldsymbol{\mu}\\ \boldsymbol{1}\\ \boldsymbol{0}\end{bmatrix}.$$

The optimal value to (4.36) is thus a tight upper bound on the MAD of $\boldsymbol{z}$ and at least as good as bound (4.31). With the optimal value to (4.36), one can construct an ambiguity set similar to (4.32) and analogously, build up the worst-case expectation corresponding to it. This MAD bound and the one of (4.31) are identical if, for example, $\mu_i=\mu_j$ and $d_i=d_j$ for all $i\neq j$.

In this way, the method of Wiesemann et al. (2014) can be used to enhance our method for aggregated random variables. Building up the upper bound on the MAD of $\boldsymbol{a}^T\boldsymbol{z}$ via problem (4.36) is preferable to bound (4.31) if the need to solve the optimization problem is not burdensome.

Independence - $3^{n_{\boldsymbol{z}}}$ terms in a pre-processing step. As a last bound, we note that the function $|\boldsymbol{a}^T\boldsymbol{z}-\boldsymbol{a}^T\boldsymbol{\mu}|$, whose expectation is equal to $M(\boldsymbol{a}^T\boldsymbol{z})$ is convex in $\boldsymbol{z}$. This means that its worst-case expectation can be computed using Ben-Tal and Hochman (1972):

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left|y-\boldsymbol{a}^T\boldsymbol{\mu}\right| &\leq \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left|y-\boldsymbol{a}^T\boldsymbol{\mu}\right| \\
&= \sum_{\boldsymbol{\alpha}\in\{1,2,3\}^{n_{\boldsymbol{z}}}}\prod_{i=1}^{n_{\boldsymbol{z}}} p^i_{\alpha_i}|\boldsymbol{a}^T\boldsymbol{z}(\boldsymbol{\alpha})-\boldsymbol{a}^T\boldsymbol{\mu}|,
\end{aligned} \tag{4.37}$$

with $\boldsymbol{z}(\boldsymbol{\alpha}) = (\tau^1_{\alpha_1}, \ldots, \tau^{n_{\boldsymbol{z}}}_{\alpha_{n_{\boldsymbol{z}}}})$ as in (4.15). This gives rise to another ambiguity set, which can be constructed analogously to (4.32). The computation in (4.37) can be conducted before the optimization problem is set up. Therefore, optimization problem involving constraint $f(\boldsymbol{x}, \boldsymbol{a}^T\boldsymbol{z})$ is easier than would be the case if formula (4.20) of Section 4.3.1 were used.

4.4.3 Vector $\boldsymbol{a}(\boldsymbol{x})$ depends on $\boldsymbol{x}$.

We assume that $\boldsymbol{a}(\boldsymbol{x})$ is a linear vector-valued function of $\boldsymbol{x}$ and the function whose worst-case expectation we seek is $f(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z})$, where $f(\cdot)$ is convex and $y(\boldsymbol{x}) = \boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z}$. This assumption also holds for the inventory problem of Section 4.3.3.

Example 4.2 *Using linear decision rules $q_{t+1}(\boldsymbol{z}^t) = q_{t+1,0} + \sum_{j=1}^t q_{t+1,j}z_j$ for the ordering decisions, the state of inventory at time $t+1$ is*

$$x_{t+1} = x_1 + \sum_{s=1}^{t}\left(q_{t,0} + \sum_{j=1}^{t} q_{t,j}z_j\right) - \sum_{j=1}^{t} z_t = x_1 + \sum_{s=1}^{t} q_{t,0} + \sum_{s=1}^{t}\left(\sum_{j=s+1}^{t} q_{j,s} - 1\right) z_s.$$

Therefore, in each time period the aggregated random variable is $\sum_{s=1}^t \left(\sum_{j=s+1}^t q_{j,s} - 1\right) z_s$, which indeed depends on the decision variables. ■

Similarly to the previous case, one can consider the worst-case expectation of $f(y(\boldsymbol{x}))$ where $y(\boldsymbol{x}) = \boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z}$, which for the set (4.32) is:

$$\begin{aligned}
\sup_{\mathbb{P}_{y(\boldsymbol{x})}\in\mathcal{P}_{y(\boldsymbol{x})}} \mathbb{E}_{\mathbb{P}_{y(\boldsymbol{x})}} f(y(\boldsymbol{x})) \quad \leq \quad & \frac{|\boldsymbol{a}(\boldsymbol{x})|^T\boldsymbol{d}}{2(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{\mu} + \|\boldsymbol{a}(\boldsymbol{x})\|_1)} f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + \\
& + \left(1 - \frac{|\boldsymbol{a}(\boldsymbol{x})|^T\boldsymbol{d}}{2(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{\mu} + \|\boldsymbol{a}(\boldsymbol{x})\|_1)} - \frac{|\boldsymbol{a}(\boldsymbol{x})|^T\boldsymbol{d}}{2(\|\boldsymbol{a}(\boldsymbol{x})\|_1 - \boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{\mu})}\right) f(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{\mu}) \\
& + \frac{|\boldsymbol{a}(\boldsymbol{x})|^T\boldsymbol{d}}{2(\|\boldsymbol{a}\|_1 - \boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{\mu})} f(\|\boldsymbol{a}(\boldsymbol{x})\|_1) \qquad (4.38)
\end{aligned}$$

This expression is not neccessarily convex in $\boldsymbol{x}$. However, for the important special case where $\boldsymbol{\mu} = 0$, satisfied for example, if $\boldsymbol{z}$ is considered to be a distortion with expected value 0 such as the white noise in signal processing, (4.38) can be bounded as follows:

$$\begin{aligned}
\sup_{\mathbb{P}_{y(\boldsymbol{x})}\in\mathcal{P}^{\boldsymbol{d}}_{y}(\boldsymbol{x})} \mathbb{E}_{\mathbb{P}_{y(\boldsymbol{x})}} f(y(\boldsymbol{x})) \quad \leq \quad & \left(1 - \frac{\sum_i |a_i(\boldsymbol{x})|d_i}{\|\boldsymbol{a}(\boldsymbol{x})\|_1}\right) f(0) + \frac{\sum_i |a_i(\boldsymbol{x})|d_i}{2\|a(\boldsymbol{x})\|_1} \left(f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + f(\|\boldsymbol{a}(\boldsymbol{x})\|_1)\right) \\
= \quad & \left(\frac{\sum_i |a_i(\boldsymbol{x})|(1-d_i)}{\|\boldsymbol{a}(\boldsymbol{x})\|_1}\right) f(0) + \frac{\sum_i |a_i(\boldsymbol{x})|d_i}{2\|\boldsymbol{a}(\boldsymbol{x})\|_1} \left(f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + f(\|\boldsymbol{a}(\boldsymbol{x})\|_1)\right) \\
\leq \quad & \max_i (1-d_i) f(0) + \max_i d_i \left(\frac{1}{2} f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + \frac{1}{2} f(\|\boldsymbol{a}(\boldsymbol{x})\|_1)\right). \qquad (4.39)
\end{aligned}$$

Quality of the bound (4.39) depends now on the dispersion of the MADs d_i - if they are equal, the second inequality is tight. Tractability of (4.39) depends on convexity

of the sum

$$f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + f(\|\boldsymbol{a}(\boldsymbol{x})\|_1), \tag{4.40}$$

which turns out to be the case, as the following proposition shows.

Proposition 4.2 *For affine $\boldsymbol{a}(\boldsymbol{x})$ and convex $f : \mathbb{R} \to \mathbb{R}$, the function $f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1)+ f(\|\boldsymbol{a}(\boldsymbol{x})\|_1)$ is convex.*

Proof. Define $g(t) = f(t) + f(-t)$ for $t \in \mathbb{R}_+$ and $h(\boldsymbol{x}) = \|\boldsymbol{a}(\boldsymbol{x})\|_1$. Then we have that:

$$f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + f(\|\boldsymbol{a}(\boldsymbol{x})\|_1) = g(h(\boldsymbol{x})).$$

The function $g(h(\boldsymbol{x}))$ is convex if $g(t)$ is convex and nondecreasing and $h(\boldsymbol{x})$ is convex. Convexity of $h(\boldsymbol{x})$ is clear as it is a norm of an affine function of $\boldsymbol{x}$. Also, convexity of $g(\cdot)$ follows from convexity of $f(\cdot)$. We need to show that $g(\cdot)$ is nondecreasing, i.e., that

$$g(t + \alpha) \geq g(t), \quad \forall t \geq 0, \alpha \geq 0.$$

Consider the subgradients $v_1 \in \partial f(-t)$ and $v_2 \in \partial f(t)$. By properties of subgradients we have that

$$v_1 \leq \frac{f(t) - f(-t)}{2t} \leq v_2.$$

From this, it follows that for $\alpha \geq 0$:

$$\begin{aligned} g(t + \alpha) &= f(-t - \alpha) + f(t + \alpha) \\ &\geq f(-t) + \sup_{v \in \partial f(-t)} (-\alpha v) + f(t) + \sup_{v \in \partial f(t)} (\alpha v) \\ &\geq f(-t) + (-\alpha v_1) + f(t) + (\alpha v_2) \\ &= g(t) + \alpha(v_2 - v_1) \\ &\geq g(t). \end{aligned}$$

■

With respect to the efficient implementation, it can be added that due to $f(-t)+f(t)$ being nondecreasing on the nonnegative ray we have:

$$f(-\|\boldsymbol{a}(\boldsymbol{x})\|_1) + f(\|\boldsymbol{a}(\boldsymbol{x})\|_1) \leq 0 \quad \Leftrightarrow \quad \begin{cases} f(-\mathbf{1}^T\boldsymbol{w}) + f(\mathbf{1}^T\boldsymbol{w}) \leq 0 \\ \boldsymbol{w} \geq \boldsymbol{a}(\boldsymbol{x}) \\ \boldsymbol{w} \geq -\boldsymbol{a}(\boldsymbol{x}), \end{cases}$$

where $\boldsymbol{w}$ is an additional analysis variable.

4.4.4 Inventory experiment revisited - independent demand

In this section we re-visit our inventory experiment using the results for aggregated random vectors of the previous section. The inventory experiment can be studied in this way, since the objective function is:

$$\sum_{t=1}^{T} \left(c_t q_t(\boldsymbol{z}^{t-1}) + h_t \max\left\{ x_{t+1}(\boldsymbol{z}^t), 0 \right\} + p_t \max\left\{ -x_{t+1}(\boldsymbol{z}^t), 0 \right\} \right),$$

where we dropped the first term as we assumed in the numerical experiment that the salvage value is zero, i.e., $s = 0$. The objective consists of T terms

$$f_t(\boldsymbol{q}, \boldsymbol{z}) = c_t q_t(\boldsymbol{z}^{t-1}) + h_t \max\left\{ x_{t+1}(\boldsymbol{z}^t), 0 \right\} + p_t \max\left\{ -x_{t+1}(\boldsymbol{z}^t), 0 \right\}, \tag{4.41}$$

that depend on the state of inventory x_{t+1} each. Therefore, in line with Examples 4.1 and 4.2, we can use our results for aggregated random variables to build worst-case expectations of $f_t(\boldsymbol{q}, \boldsymbol{z})$ and to use them in the optimization problem.

Since the methods of this section are also aimed at an alleviation of the independence assumption and the piecewise linear objective function is tractable using the results of Wiesemann et al. (2014), we compare our solutions also to Wiesemann et al. (2014), whose results do not rely on the independence assumption either. We note here that to use their results to obtain an exact reformulation, we need to formulate the sum-of-maximums objective function as a maximum of linear functions which leads to 2^T terms. For the case where the independence assumption is satisfied, we refer the reader to Appendix 4.C where the difference between their and our reformulations is illustrated.

In the following, we consider seven solutions to the same 50 instances as in Sections 4.3.3 and 4.3.4. The first four solutions consider the ordering decisions to be static:

- S-1: using Ben-Tal and Hochman (1972) to evaluate the true worst-case expectation of the objective with 3^T terms in the problem formulation.
- S-2: using the aggregated random variables $y = \mathbf{1}^T \boldsymbol{z}$ for which an upper bound on the MAD is computed under the independence assumption with (4.37) (T terms in the problem formulation).
- S-3: using the methodology of Wiesemann et al. (2014) (2^T terms in the problem formulation).
- S-4: using ambiguity set (4.32) without the independence assumption on z_i's ($3T$ terms in the problem formulation).

The other three solutions consider the ordering decisions to be linear functions of past demand:

Table 4.5 – Results of the inventory management with aggregated random vectors - worst-case costs and ranges for the expectation of the objective over $\mathcal{P}_{(\mu,d,\beta)}$ (with independence assumption in computing the bounds). The lower bounds of the intervals are obtained *ex-post* using Ben-Tal and Hochman (1972) results, after the solutions are found. All numbers are averages.

β	Cost expectation range						
	S-1	S-2	S-3	S-4	LDR-1	LDR-2	LDR-3
0.25	[1093,1175]	[1124,1205]	[1131,1190]	[1131,1190]	[1004,1049]	[1058,1093]	[1058,1092]
0.5	[1038,1175]	[1065,1205]	[1061,1190]	[1061,1190]	[970,1049]	[1029,1093]	[1028,1092]
0.75	[1079,1175]	[1117,1205]	[1106,1190]	[1106,1190]	[994,1049]	[1047,1093]	[1047,1092]

- LDR-1: using Ben-Tal and Hochman (1972) to evaluate the true worst-case expectation of the objective with 3^T terms.
- LDR-2: using methodology of Wiesemann et al. (2014), (2^T terms in the problem formulation).
- LDR-3: using approximation (4.39) to obtain upper bound on the worst-case expectation of the objective ($3T$ terms in the problem formulation).

Table 4.5 presents the worst-case and best-case expectations for all 7 solutions, computed under the assumption of independence of z_i (in the same way as in the main experiment). For the solutions with static decisions we can see that the new solutions S-2 and S-4 based on the aggregation technique are only slightly worse than the original solution based on Ben-Tal and Hochman (1972) bound with 3^T terms. For example, their worst-case expectations are 1205 and 1190, respectively, whereas solution S-1 yields 1175, which makes them only 2% worse than the exact formulation.

For the solutions with linear decision rules, we see that the expectation intervals overlap for $\beta = 0.5$ and $\beta = 0.75$. The worst-case expectation of the new LDR-3 solution is 1092, whereas for the old S-1 solution it is 1049. That is, the new solution is about 4% worse than the exact formulation LDR-1.

We observe that the intervals obtained by our aggregated solutions are very similar to the ones obtained by the methodology of Wiesemann et al. (2014), compare solutions S-3 vs S-4 and LDR-2 vs LDR-3. Solutions S-3 and S-4 are identical for all instances since our aggregation technique is exact in this case, just as the method of Wiesemann et al. (2014) - this is because in each case, the mean of the uncertain demand is in the middle of the support and the proportion of the MADs of z_t's to the support intervals' widths is the same for all t, see the description of the setting in Section 4.3.3. However, the similarity of the intervals for LDR-2 and LDR-3 comes as a surprise since bound (4.39) is just an approximation, whereas the method of Wiesemann et al. (2014) is exact and involves 2^T terms in the problem formulation (their results rely

Table 4.6 – Simulation results for the inventory problem with aggregation technique. Numbers in brackets denote the % change compared to the S-1/LDR-1 solution, respectively.

Objective type	Sample	Cost						
		S-1	S-2	S-3	S-4	LDR-1	LDR-2	LDR-3
Mean	Uniform	1088	1182 (+8.6%)	1109 (+1.9%)	1109 (+1.9%)	994	1051 (+5.7%)	1049 (+5.5%)
Standard deviation	Uniform	317	325 (+2.5%)	321 (+1.36%)	321 (+1.36%)	259	252 (−2.7%)	251 (−3.1%)
Mean	(μ, d)	1094	1124 (+2.7%)	1115(+1.9%)	1115(+1.9%)	1003	1056 (+ 5.5%)	1018 (+1.5%)
Standard deviation	(μ, d)	325	333 (+2.5%)	328 (+0.9%)	328 (+0.9%)	265	257 (−3.0%)	224 (−15.5%)

on formulating the objective as a maximum over a finite number of affine functions, which for our sum-of-max objective requires 2^T terms to consider all cases).

Table 4.6 presents the results of simulation in the same setting as in Table 4.3, with demands from different periods being independent. As we can see, the new S-2, S-4 and LDR-3 solutions perform slightly worse on average than the exact S-1 and LDR-1 solutions. An important observation, however, is that the solution LDR-2 based on the rather conservative MAD bound performs better than the exact S-1 solution which utilizes static decisions.

We can see that the solutions S-3 (Wiesemann et al. 2014) and S-4 (using (4.31)) are indeed the same since the simulation results are identical for both. Comparing the solutions LDR-2 (Wiesemann et al. 2014) and LDR-3 (using (4.39)) we see that despite the similarity of intervals in Table 4.5 the solutions do differ, since the obtained results are not the same. The large difference between the two for the (μ, d) sample (means 1056 and 1018, respectively) is a sample-specific issue and, having repeated the experiment in multiple samples, we conclude that the two solutions give very similar results on average.

To conclude on the results of this section, we can say that for problems with aggregated random variables $\boldsymbol{a}^T\boldsymbol{z}$ our approach is preferable if there is no assumption of independence of $\boldsymbol{z}$ and (i) the need to compute the exact worst-case MAD of $\boldsymbol{a}^T\boldsymbol{z}$ via solving problem (4.36) would be too burdensome, or (ii) the complexity of the function $f(\boldsymbol{x}, \boldsymbol{a}^T\boldsymbol{z})$ is not tractable in the sense of requirements of Wiesemann et al. (2014). In other cases, it is preferable either to estimate the exact worst-case MAD of $\boldsymbol{a}^T\boldsymbol{z}$ via (4.36) or to apply the results of Wiesemann et al. (2014) directly. For the case of $\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z}$, our method is preferable if there is no assumption of independence of $\boldsymbol{z}$ and the complexity of the function $f(\boldsymbol{a}(\boldsymbol{x})^T\boldsymbol{z})$ is not tractable in the sense of requirements of Wiesemann et al. (2014). Otherwise, the approach of Wiesemann et al. (2014) is preferred.

4.4.5 Inventory experiment revisited - dependent demand

In the previous subsection the solutions were evaluated using demand samples in which the demands from subsequent periods were independent. One may ask: how do the solutions perform when the realized demand sample exhibits some dependence pattern?

To investigate this, we run an experiment where the demands $\boldsymbol{z}$ are sampled using copulas that couple a multivariate distribution function to its marginal distributions. Separating the dependence structure between random variables from their marginal distributions makes them a premier tool for simulating correlated random variables when particular marginal distributions are desired (Sklar 1959). In our case, we want the marginal distributions to come from our (μ, d) sample (results for the uniform marginals are very similar).

In our experiment, we use the T-dimensional Gaussian copula[3] and assume that the dependence between the z_i's follows an autocorrelative pattern where the correlations between the random variables used in the copula from periods t_1 and t_2 is equal to $\rho^{|t_1-t_2|}$, where $\rho \in \{0.1, 0.2, \ldots, 0.9\}$, that is, for the copula sampling we use the correlation matrix:

$$\begin{bmatrix} 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^1 & 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 \\ \vdots & & \ddots & & & \vdots \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}.$$

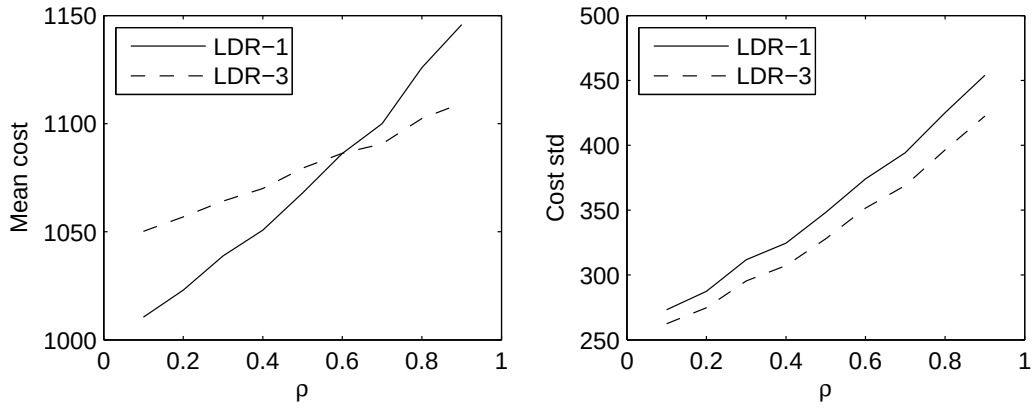
For conciseness, we focus only on the LDR-1 and LDR-3 solutions under (μ, d) sample.

Figure 4.3 presents the results, with the respective means and standard deviations plotted against the correlation strength ρ. The left panel shows that as the degree of correlation among the demands increases with ρ (on the horizontal axis), the mean costs obtained by solution LDR-3 approaches the one of LDR-1 and eventually becomes smaller, with the crossing point approximately around $\rho = 0.6$. In the right panel we see that the standard deviation of the costs obtained by LDR-3 is is smaller than the one of LDR-1 for all values of ρ.

These results provide a strong argument that the LDR-3 solution, though being constructed on the basis of approximation (4.39) and not assuming the independence of z_i's might be better than LDR-1 solution based on the full (μ, d) information in situations when the realized demand exhibits dependence among its components from different periods. More generally, this indicates that the solutions based on

[3]We use the MATLAB function copularnd(). As this function simulates only the CDFs of the marginal distributions, we need to convert them into the respective uniform and (μ, d) sample by inverting their distribution functions.

Figure 4.3 – Results of the simulation experiment for inventory solutions LDR-1 and LDR-3 with dependent (μ, d) demand sample. The left panel involves results on the means and the right panel on the standard deviations.



aggregated random vectors, while being less computationally burdensome than the 'exact solutions' can yield better performance when the true random variables deviate from the independence assumption.

4.5 Safe approximations of chance constraints

4.5.1 Introduction

In this section we show how the results of Ben-Tal and Hochman (1972) can be used to construct safe tractable approximations of scalar chance constraints:

$$\mathbb{P}\left(\boldsymbol{a}^T(\boldsymbol{z})\boldsymbol{x} > \boldsymbol{b}(\boldsymbol{z})\right) \leq \epsilon, \quad \forall \mathbb{P} \in \mathcal{P}_{(\mu,d)}, \text{ where } [\boldsymbol{a}(\boldsymbol{z}); \boldsymbol{b}(\boldsymbol{z})] = [\boldsymbol{a}^0; \boldsymbol{b}^0] + \sum_{i=1}^{n_{\boldsymbol{z}}} z_i [\boldsymbol{a}_i^0; \boldsymbol{b}_i^0]. \tag{4.42}$$

Without loss of generality we assume that the components z_i have a support contained in $[-1, 1]$ and mean 0:

$$\mathcal{P}_{(\mu,d)} = \{\mathbb{P} : \quad \text{supp}(z_i) \subseteq [-1,1], \quad \mathbb{E}_{\mathbb{P}} z_i = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d_i, \quad i = 1, \ldots, n_{\boldsymbol{z}}, \quad z_i \perp\!\!\!\perp z_j, \quad \forall i \neq j\}.$$

To construct the safe tractable approximations, we use the mathematical framework of Ben-Tal et al. (2009). In this framework, the key step consists of bounding from above the *moment-generating function* of $z_i, i = 1, \ldots, n_{\boldsymbol{z}}$:

$$\mathbb{E}_{\mathbb{P}} \exp(w z_i) = \int \exp(w z_i) d\mathbb{P}_i(z_i)$$

and then using the resulting bound in combination with the Markov inequality to obtain upper bounds on the probability $\mathbb{P}(\boldsymbol{a}^T(\boldsymbol{z})\boldsymbol{x} > \boldsymbol{b}(\boldsymbol{z}))$ - often referred to as the *Bernstein approximation*.

A strong motivation for using the ambiguity set $\mathcal{P}_{(\mu,d)}$ is due to the fact that a tight explicit bound on $\mathbb{E}_{\mathbb{P}} \exp(\boldsymbol{w}^T \boldsymbol{z})$ is obtained easily in this setting by the Ben-Tal and Hochman (1972) results described in Section 2. Indeed, due to the independence of $z_1, \ldots, z_{n_{\boldsymbol{z}}}$ we have:

$$\begin{aligned}
\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp(\boldsymbol{z}^T \boldsymbol{w}) &= \prod_{i=1}^{n_{\boldsymbol{z}}} \sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp(z_i w_i) \\
&= \prod_{i=1}^{n_{\boldsymbol{z}}} \left(\frac{d_i}{2} \exp(-w_i) + \frac{d_i}{2} \exp(w_i) + (1 - d_i) \exp(0) \right) \\
&= \prod_{i=1}^{n_{\boldsymbol{z}}} \left(d_i \cosh(w_i) + 1 - d_i \right). \qquad (4.43)
\end{aligned}$$

The worst-case expectation is evaluated separately for each component of $\boldsymbol{z}$, avoiding the computational burden of summation of $3^{n_{\boldsymbol{z}}}$ terms as in (4.14). In Appendix 4.C we show that in the setting of Wiesemann et al. (2014) without independence of z_i's, obtaining the tight upper bound on $\exp(\boldsymbol{w}^T \boldsymbol{z})$ requires solving an optimization problem involving an uncertain constraint on a convex function. This requires an exponential number of constraints for an exact reformulation.

4.5.2 Safe approximations - results

We now show how (4.43) can be used to obtain safe approximations of (4.42). First, we present two simple safe approximations in order of increasing tightness. Later, we show that the (μ, d) information is particularly suitable for obtaining even tighter safe approximations, based on the exponential polynomials.

The first approximation requires the use of Theorem 2.4.4 of Ben-Tal et al. (2009), repeated in Appendix 4.B.1.

Theorem 4.1 *If for a given vector $\boldsymbol{x}$ there exist $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n_{\boldsymbol{z}}+1}$ such that $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})$ satisfies the constraint system*

$$\begin{cases}
(\boldsymbol{a}^i)^T \boldsymbol{x} - b_i = u_i + v_i, 0 \leq i \leq n_{\boldsymbol{z}} \\
u_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} |u_i| \leq 0 \\
v_0 + \sqrt{2 \log(1/\epsilon)} \sqrt{\sum_{i=1}^{n_{\boldsymbol{z}}} \sigma_i^2 v_i^2} \leq 0,
\end{cases} \qquad (4.44)$$

where

$$\sigma_i = \sup_{t \in \mathbb{R}} \sqrt{\frac{2 \log\left(d_i \cosh(t) + 1 - d_i\right)}{t^2}}, \qquad (4.45)$$

then $\boldsymbol{x}$ is feasible to (4.42), that is, constraint system (4.44) is a safe approximation of (4.42). Moreover, (4.44) is the robust counterpart of the following robust constraint

$$\boldsymbol{a}^T(\boldsymbol{z})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{U}, \text{ where } [\boldsymbol{a}(\boldsymbol{z}); \boldsymbol{b}(\boldsymbol{z})] = [\boldsymbol{a}^0; \boldsymbol{b}^0] + \sum_{i=1}^{n_{\boldsymbol{z}}} z_i [\boldsymbol{a}_i^0; \boldsymbol{b}_i^0], \quad (4.46)$$

and

$$\mathcal{U} = \left\{ \boldsymbol{z} \in \mathbb{R}^{n_{\boldsymbol{z}}} : \quad \sqrt{\sum_{i=1}^{n_{\boldsymbol{z}}} \frac{z_i^2}{\sigma_i^2}} \leq \sqrt{2\log(1/\epsilon)}, \quad -1 \leq z_i \leq 1, \quad i = 1, \ldots, n_{\boldsymbol{z}} \right\}.$$

Proof. The proof follows the steps leading to Theorem 2.4.4 from Ben-Tal et al. (2009). First, we need to find scalars $\mu_i^-, \mu_i^+, \sigma_i$, where $i = 1, \ldots, n_{\boldsymbol{z}}$ such that:

$$\int_{-1}^{1} \exp(tz_i) d\mathbb{P}_i(z_i) \leq \exp\left(\max\{\mu_i^-, \mu_i^+\} + \frac{\sigma_i^2}{2} \right), \quad \forall t \in \mathbb{R}, \quad \forall \mathbb{P} \in \mathcal{P}_{(\mu,d)}.$$

By (4.43) we have $\sup\limits_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \left\{ \int\limits_{-1}^{1} \exp(tz_i) d\mathbb{P}_i(z_i) \right\} = d_i \cosh(t) + 1 - d_i$. Thus, for each i we need to find $\mu_i^-, \mu_i^+, \sigma_i$ such that:

$$d_i \cosh(t) + 1 - d_i \leq \exp\left(\max\{\mu_i^- t, \mu_i^+ t\} + \frac{\sigma_i^2 t^2}{2} \right), \quad \forall t \in \mathbb{R}.$$

Setting $\mu_i^- = \mu_i^+ = 0$, we then need σ_i such that

$$\begin{aligned} & d_i \cosh(t) + 1 - d_i \leq \exp\left(\frac{\sigma_i^2 t^2}{2} \right), \quad \forall t \in \mathbb{R} \quad \Leftrightarrow \\ \Leftrightarrow \quad & \sigma_i^2 \geq g_i(t) = \frac{2}{t^2} \log\left(d_i \cosh(t) + 1 - d_i \right), \quad \forall t \in \mathbb{R}. \end{aligned}$$

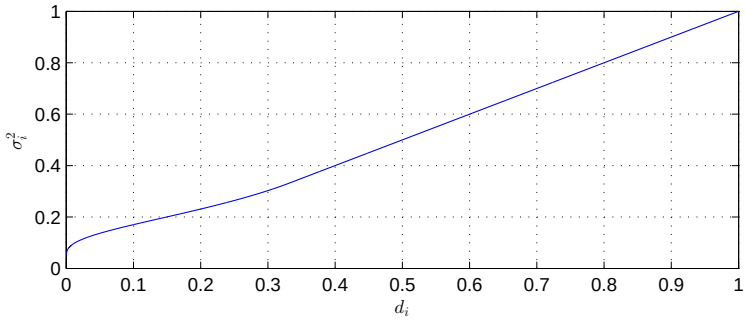
Thus, we look for the maximum value of $g_i(t)$ over the real axis. From the definition of $g_i(t)$ we know that it is finite, nonnegative, and differentiable everywhere except for 0. By de l'Hôpital rule we have that $\lim\limits_{t \to 0} g_i(t) = d_i$. It holds that $\lim\limits_{t \to \pm\infty} g_i(t) = 0$. The value of σ_i can be obtained by means of a numerical analysis. Figure 4.4 presents the plot of σ_i^2 as a function of d_i.

From here, by inserting the values $\mu_i^- = \mu_i^+ = 0$, and σ_i into Theorem 2.4.4 of Ben-Tal et al. (2009) (see Appendix 4.B.1), we obtain that robust constraint (4.46) with $\mathcal{U}$ defined as above, is a safe tractable approximation of chance constraint (4.42). By the same theorem, it holds that (4.44) is the robust counterpart of (4.46). ■

Constraint system (4.44) involves only linear and second-order conic constraints, making it highly tractable even for large-dimensional problems.

The second safe approximation is tighter and relies on the somewhat more complicated mathematical machinery of Ben-Tal et al. (2009).

Figure 4.4 – Plot of σ_i^2 as a function of d_i.



Theorem 4.2 *If there exists $\alpha > 0$ such that $(\boldsymbol{x}, \alpha)$ satisfies the constraint*

$$(\boldsymbol{a}^0)^T\boldsymbol{x} - b_0 + \alpha \log\left(\sum_{i=1}^{n_{\boldsymbol{z}}}\left(d_i \cosh\left(\frac{(\boldsymbol{a}^i)^T\boldsymbol{x} - b_i}{\alpha}\right) + 1 - d_i\right)\right) + \alpha\log(1/\epsilon) \leq 0, \quad (4.47)$$

then $\boldsymbol{x}$ satisfies constraint (4.42). That, is (4.47) is a safe approximation of (4.42).

Proof. See Appendix 4.B.2. ■

Similar to Theorem 4.1, one can construct an explicit convex uncertainty set $\mathcal{U}$ for which (4.47) is the robust counterpart of (4.46) corresponding to $\mathcal{U}$. Constraint (4.47) is convex in $(\boldsymbol{x}, \alpha)$, being a sum of a linear function and $n_{\boldsymbol{z}}$ perspective functions of the convex log-sum-exp function, see Boyd and Vandenberghe (2004). For that reason, it can be handled with convex optimization algorithms such as Interior Point Methods.

4.5.3 Towards better safe approximations - exponential polynomials

Ben-Tal et al. (2009) discuss the fact that the bounds obtained using a single exponential function can still be improved by, instead of the moment-generation function, constructing the worst-case expectation of exponential polynomials:

$$\gamma(s) = \sum_{\nu=0}^{L} c_\nu \exp\{\omega_\nu s\}, \quad (4.48)$$

to bound the probability of constraint violation, where $c_\nu, \omega_\nu, \nu = 0, \ldots, L$ are complex numbers and

$$\gamma(\cdot) \text{ is convex and nondecreasing, } \gamma(s) \geq 0, \quad \gamma(0) \geq 0, \quad \gamma(s) \to 0, \quad s \to -\infty. \quad (4.49)$$

The worst-case expectation of the exponential polynomial $\gamma(s)$, similar to the worst-case expectation of the moment-generating function (4.43), can then be used to obtain better upper bounds on $\mathbb{P}\left(\boldsymbol{a}^T(\boldsymbol{z})\boldsymbol{x} > \boldsymbol{b}(\boldsymbol{z})\right)$. In fact, the bound found in Theorem 4.2

is obtained using a special case of (4.48), where $L = 0$, $c_0 = \omega_0 = 1$. The difficulty of using general polynomials (4.48) lies in the (un)availability of tight, computationally tractable upper bounding function $\Psi(\boldsymbol{w})$ on (4.48):

$$\mathbb{E}_{\mathbb{P}}\gamma\left(w_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} w_i z_i\right) \leq \Psi(\boldsymbol{w}), \quad \forall \mathbb{P} \in \mathcal{P}.$$

In the following, we show that under (μ, d) information, the result of Ben-Tal and Hochman (1972) can be easily applied in this case as well. Indeed, the corresponding supremum over $\mathcal{P}_{(\mu,d)}$ is given by:

$$\begin{aligned} \Psi(\boldsymbol{w}) &= \sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}}\gamma\left(w_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} z_i w_i\right) \\ &= \sum_{\nu=0}^{L} c_\nu \exp\{\omega_\nu w_0\} \prod_{i=1}^{n_{\boldsymbol{z}}} \left(d_i \sinh(\omega_\nu w_i) + 1 - d_i\right). \end{aligned} \tag{4.50}$$

Now, we can use Proposition 4.3.1 from Ben-Tal et al. (2009) to obtain the following result.

Theorem 4.3 *Consider an exponential polynomial $\gamma(s)$ satisfying (4.49), the corresponding function $\Psi(\boldsymbol{w})$ and the set Γ_ϵ such that:*

$$\Gamma_\epsilon = \{\boldsymbol{x} : \quad \exists \alpha > 0 : \quad \Psi(\alpha \boldsymbol{w}) \leq \epsilon\}, \qquad w_i = (\boldsymbol{a}^i)^T \boldsymbol{x} - b_i, \quad i = 1, \ldots, n_{\boldsymbol{z}}. \tag{4.51}$$

Then, any $\boldsymbol{x} \in \mathrm{cl}\Gamma_\epsilon$ is also feasible for the chance constraint (4.42).

Proof. See Appendix 4.B.3. ■

It is also important to note that constraint (4.51) is convex representable in $(\boldsymbol{x}, \alpha)$. Theorem 4.3 extends the results of Ben-Tal et al. (2009), which provides a safe approximation using only known supports and means of the components z_i.

4.5.4 Safe tractable approximations - simple experiment

We illustrate now the differences between (i) the power of the three approximations of the previous sections, and (ii) knowing and not knowing the MAD. We consider here the following problem from Section 4.3.6.2 of Ben-Tal et al. (2009):

$$\begin{aligned} \max_{x_0} \quad & x_0 \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}_{(\mu,d)}} \mathbb{P}\left(x_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} x_i z_i > 0\right) \leq \epsilon \\ & x_i = 1, \quad i = 1, \ldots, n_{\boldsymbol{z}}. \end{aligned} \tag{4.52}$$

We solve this problem using all three safe tractable approximations of the chance constraint, for two different cases:

Table 4.7 – Maximum values of x_0 in problem (4.52), depending on the safe tractable approximation used, probability bound, and the assumptions on the knowledge about d.

ϵ		Maximum x_0		
	Safe approximation	Theorem 1	Theorem 2	Theorem 3
10^{-1}	Unknown d	-24.28	-24.21	-20.43
	$d = 0.5$	-17.16	-17.14	-14.48
10^{-2}	Unknown d	-34.34	-34.13	-30.55
	$d = 0.5$	-24.27	-24.20	-21.69
10^{-3}	Unknown d	-42.05	-41.67	-38.34
	$d = 0.5$	-29.73	-29.60	-27.25

- no information about d - which corresponds to setting $d_i = 1, i = 1, \ldots, n_{\boldsymbol{z}}$ (the largest possible value for d_i, see Remark 4.1, about the Edmundson-Madansky bound when d is maximum possible),
- knowing that $d_i = d = 0.5, i = 1, \ldots, n_{\boldsymbol{z}}$.

We consider three probability levels: $\epsilon \in \{10^{-1}, 10^{-2}, 10^{-3}\}$ and $n_{\boldsymbol{z}} = 128$. Whereas safe approximations corresponding to Theorems 1 and 2 are well-defined by the theorems, we need to choose the exponential polynomial used in the approximation of Theorem 3. As Ben-Tal et al. (2009), we use the polynomial

$$\gamma_{d,T}(s) = \exp(s)\chi_{c^*}(s),$$

where

$$\chi_{c^*}(s) = \sum_{\nu=0}^{d} \left(c_\nu^* \exp(\imath\pi\nu s/T) + \overline{c_\nu^*} \exp(-\imath\pi\nu s/T)\right)$$

is an optimal solution of the best uniform approximation problem:

$$\boldsymbol{c}^* \in \operatorname{argmin}\left\{ \max_{-T\le s\le T} |\exp(s)\chi_{\boldsymbol{c}}(s) - \max\{1+s, 0\}| : \quad 0 \le \chi_{\boldsymbol{c}}(s) \le \chi_{\boldsymbol{c}}(0) = 1, \quad \forall s \in \mathbb{R} \right\}$$

and $\exp(s)\chi_{\boldsymbol{c}}(s)$ is convex nondecreasing on $[-T, T]$, with parameter values $d = 11$ ('degree of approximation' of the function $\max\{1+s, 0\}$), $T = 8$ ('window width' on which the function $\max\{1+s, 0\}$ is approximated).

Table 4.7 presents the results. First, for all safe approximations and all security levels, one can observe a substantial value of having the information about parameters d_i. For example, for $\epsilon = 0.01$ and safe approximation according to Theorem 3, the optimal solution obtained without knowledge of d is -30.55, whereas the corresponding number for known $d = 0.5$ is -21.69. A similar pattern can be observed for other values of ϵ and other approximations.

Secondly, one can see the increasing power of the safe tractable approximations that use exactly the same information. For example, for $\epsilon = 10^{-3}$ and $d = 0.5$ the

subsequent optimal values are -29.73, -29.60 and -27.25. For all values of ϵ and d there is a bigger difference between the second and third tractable approximation than between the first and second.

This example illustrates the extra power due to the knowledge of d, giving a strong reason to estimate this quantity in order to obtain better chance constraint approximations. Also, the difference between the quality of safe tractable approximations of Theorems 1, 2, and 3 illustrates that the power of exponential polynomial-based approximations make them an attractive tool if the parameters a, b, μ, and d can be estimated with sufficient precision.

4.5.5 Antenna array - chance constraints

Here we consider an application of our safe tractable approximations to scalar chance constraints to an antenna design problem under implementation error uncertainty.

Antenna is a device for sending and receiving electromagnetic signals. The signal emitted by an antenna corresponds to a function called *diagram*. An antenna array is a system of several antennas whose diagram is the sum of the diagrams of the individual components. In designing the antenna array the engineer can amplify the power sent to each of the antennas so as to obtain an array whose diagram satisfies some desired properties. For more information we refer the reader to Section 3.3 of Ben-Tal et al. (2009).

In our example, the setting is as follows. There are $n = 40$ ring-shaped antennas belonging to the XY plane in $\mathbb{R}^3$. The radius of the k-th antenna is defined as k/n and the diagram $D(\varphi)$ of the antenna array is defined as a sum of diagrams $D_k(\varphi)$ of the antennas:

$$D_k(\varphi) = \frac{1}{2}\int_0^{2\pi} \cos\left(\frac{2\pi k}{40}\cos(\varphi)\cos(\gamma)\right) d\gamma, \quad k = 1, \ldots, 40.$$

The objective of the problem is to minimize the maximum of the diagram modulus in the sidelobe angle $0 \leq \varphi \leq 70°$:

$$\max_{0\leq\varphi\leq 70°} \left|\sum_{k=1}^{n} x_k D_k(\varphi)\right|,$$

where x_k are the decision variables - the amplification weights, subject to the restrictions that:

- the diagram in the interval $77° \leq \varphi \leq 90°$ is nearly uniform:

$$77° \leq \varphi \leq 90° \quad \Rightarrow \quad 0.9 \leq \sum x_k D_k(\varphi) \leq 1,$$

- the diagram in other angles is not too large:

$$\left|\sum_{k=1}^{n} x_k D_k(\varphi)\right| \leq 1, \quad 70^\circ \leq \varphi \leq 77^\circ.$$

We assume that the implementation error affects the weight of the k-th antenna in the following fashion:

$$x_k \mapsto \tilde{x}_k = (1 + z_k\rho)x_k, \quad k = 1, \ldots, n,$$

where z_k, $k = 1, \ldots, n$, are independent random variables with supports contained in $[-1, 1]$, with mean 0 and MAD equal to d:

$$\mathcal{P} = \{\mathbb{P} : \quad \text{supp}(z_i) \subset [-1, 1], \quad \mathbb{E}_{\mathbb{P}}(z_i) = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d, \quad z_i \perp\!\!\!\perp z_j, \quad \forall i \neq j\}.$$

The problem to be solved is:

$$\begin{array}{lllll}
\min\limits_{\tau, \boldsymbol{x}} & \tau & & & \\
\text{s.t.} & \mathbb{P}\left(\sum D_k(\varphi_i)\tilde{x}_k \leq \tau\right) & \geq 1-\epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 0 \leq \varphi_i \leq 70^\circ \\
& \mathbb{P}\left(\sum D_k(\varphi_i)\tilde{x}_k \geq -\tau\right) & \geq 1-\epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 0 \leq \varphi_i < 70^\circ \\
& \mathbb{P}\left(\sum D_k(\varphi_i)\tilde{x}_k \leq 1\right) & \geq 1-\epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 70^\circ \leq \varphi_i < 77^\circ \\
& \mathbb{P}\left(\sum D_k(\varphi_i)\tilde{x}_k \geq -1\right) & \geq 1-\epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 70^\circ \leq \varphi_i \leq 77^\circ \\
& \mathbb{P}\left(\sum D_k(\varphi_i)\tilde{x}_k \leq 1\right) & \geq 1-\epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 77^\circ \leq \varphi_i \leq 90^\circ \\
& \mathbb{P}\left(\sum D_k(\varphi_i)\tilde{x}_k \geq 0.9\right) & \geq 1-\epsilon, & \forall \mathbb{P} \in \mathcal{P}, & \forall 77^\circ \leq \varphi_i \leq 90^\circ,
\end{array} \tag{4.53}$$

where $\varphi_1, ..., \varphi_N$ is a 'fine grid' of equidistant points in $[0^\circ, 90^\circ]$.

In the numerical experiment we set $N = 400, d = 0.5$. The chance constraints are reformulated using the ball-box uncertainty set of Theorem 1. We solve the problem in the following settings:

- nominal solution, with $\rho = 0$ (no implementation error),
- robust solutions where both ϵ and ρ can get values in $\{0.001, 0.01, 0.05\}$.

In total, we obtain 10 solutions. For each of them we report the optimal (worst-case) objective value. Next to that, we conduct a simulation study for each solution, where the realized error magnitude $\widehat{\rho}$ can take the values in $\{0.001, 0.01, 0.05\}$. In this study, for each solution we sample 10^4 scenarios of the implementation error $\hat{\boldsymbol{z}} \in [-1, -1]^n$ and we report on:

- the percentage of samples for which at least one of the constraints of the problem (4.53) is violated,

Table 4.8 – Minimum worst-case τ^* and mean simulated values of $\widehat{\tau}$ for each of the solutions. ϵ and ρ denote the parameter values used in problem (4.53) to obtain a given solution, and $\widehat{\rho}$ denotes the error magnitude of the given sample of 10^4 implementation error vectors $\boldsymbol{z}$.

	ϵ	-	0.001			0.01			0.05		
	ρ	0	0.001	0.01	0.05	0.001	0.01	0.05	0.001	0.01	0.05
Worst-case $\tau \times 10^{-2}$	-	2.68	5.86	8.12	34.05	5.77	7.92	16.04	5.68	7.71	11.80
Average $\hat{\tau} \times 10^{-2}$	$\widehat{\rho} = 0.001$	4270	5.64	7.66	31.91	5.59	7.53	14.70	5.54	7.38	10.84
	$\widehat{\rho} = 0.01$	42706	7.03	7.78	31.91	6.96	7.65	14.70	6.91	7.51	10.85
	$\widehat{\rho} = 0.05$	213534	14.00	8.61	31.91	13.73	8.46	14.94	13.64	8.36	11.12

Table 4.9 – Empirical probabilities of violating at least one constraint. 'Violation probability (%)' denotes the percentage of simulated implementation error vectors for which at least one of the constraints of the problem (4.53) has been violated.

	ϵ	-	0.001			0.01			0.05		
	ρ	0	0.001	0.01	0.05	0.001	0.01	0.05	0.001	0.01	0.05
Violation probability (%)	$\widehat{\rho} = 0.001$	100	0.00	0.00	0.00	0.02	0.00	0.00	0.18	0.00	0.00
	$\widehat{\rho} = 0.01$	100	84.39	0.00	0.00	84.87	0.00	0.00	85.67	0.20	0.00
	$\widehat{\rho} = 0.05$	100	99.57	63.89	0.00	99.66	62.97	0.03	99.61	67.35	0.43

- the perturbed objective function value $\hat{\tau} = \max\limits_{0^\circ \leq \varphi_i \leq 70^\circ} |\sum \hat{x}_k D_k(\varphi_i)|$.

Results are given in Tables 4.8 and 4.9. The nominal solution becomes senseless already with the implementation error $\widehat{\rho} = 0.001$. At the same time, the robust solutions yield good performance even with the largest $\widehat{\rho}$, both in terms of the $\hat{\tau}$ values and the percentage of drawings for which at least one constraint is violated.

The difference between the nominal and robust solutions can be seen in Figures 4.5 and 4.6, where the diagrams are plotted for the situations (i) with no implementation error, and (ii) with a single sample of implementation error $\widehat{\rho} = 0.001$. In both cases, the solutions yield good 'desired' diagrams in the no-error case. However, in the situation with implementation error (lower panels), the robust solution still 'fits' into the desired bounds, which is completely not the case for the nominal solution.

4.6 Summary

In this paper, we have considered two types of ambiguous stochastic constraints - expected feasibility constraints and chance constraints. In contrast to previous research, which employs the variance as a dispersion measure, we use the mean absolute deviation. This allows us to use the 1972 results of Ben-Tal and Hochman (1972) on tight upper and lower bounds on the expectation of a convex function of a random variable, and thus, to provide tractable exact robust counterparts for expected feasibility constraint and to obtain safe tractable approximations of am-

Figure 4.5 – Nominal solution - diagram plots. Upper panel - situation without implementation error. Lower panel - implementation error, single trajectory.

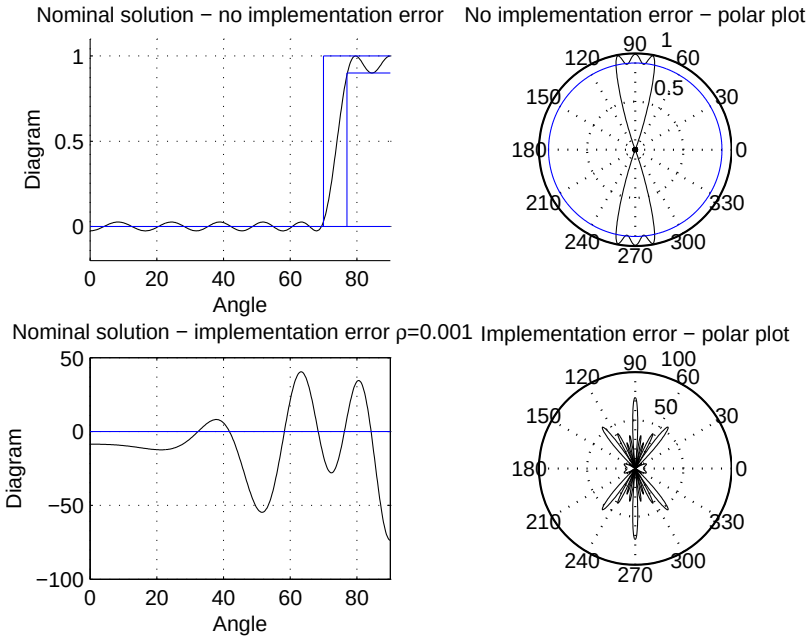
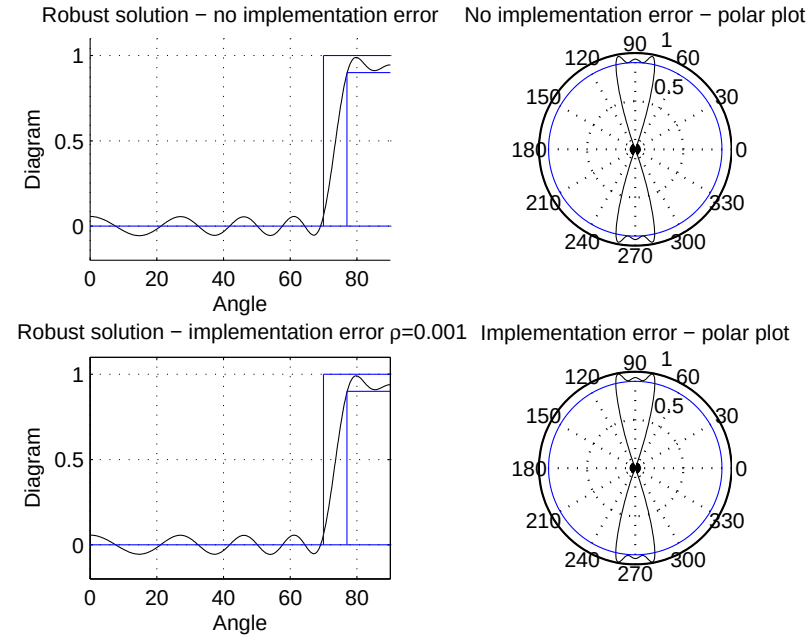


Figure 4.6 – Robust solution - diagram plots. Upper panel - situation without implementation error. Lower panel - implementation error, single trajectory.



biguous chance constraint. Numerical examples show the proposed methodology to perform well and, in particular, to offer substantial improvements in the worst-case expected performance and probabilistic guarantees on constraints' feasibility. In particular, for the worst-case expected feasibility constraints we identify an important class of functions for which we can relax the assumption of independence of random variables needed by Ben-Tal and Hochman (1972), and for which we construct highly computationally tractable approximations. Numerical experiments show that these approximations yield good practical performance and can be preferred in settings where the independence assumption on the random variables does not hold.

Appendices

4.A Estimating μ, d, and β

As the bounds on the expectation of a random variable depend on the parameters a, b, μ, d, and β, it is necessary to know or estimate these parameters, and decide 'how much information is actually available'. Here, we provide the reader with a simple procedure to achieve this.

First of all, it is necessary to verify the independence of the components of $\boldsymbol{z}$. This can be achieved using the nonparametric tests of Pinkse (1998) and Ghoudi et al. (2001). If the independence hypothesis is rejected, factorization techniques mentioned in Section 4.3.1 can be used to decompose the random variable into a combination of factors.

Assuming that the independence holds or is achieved by factorization, we operate here with the one-dimensional case for z, and the multi-dimensional case follows straightforwardly due to the independence of the components of $\boldsymbol{z}$. Appendix 4.D describes the properties of the MAD in relation to the variance and formulas for the MAD of several important classes of probability distribution.

First, we introduce estimators of μ, d, and β and discuss their asymptotic properties. Based on these results, we provide a procedure that can be used to assess whether the amount of information available is sufficient to use the results for the (μ, d) ambiguity set or the (μ, d, β) ambiguity set.

Let $z^{(1)}, \ldots, z^{(n)}$ be a random sample of the values of z. We assume the interval $[a, b]$ to be fixed by the user. As estimators for μ, d, and β we consider

- $\widehat{\mu} = \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$, the sample mean;
- $\widehat{d} = \frac{1}{n} \sum_{i=1}^{n} |z_i - \bar{z}|$, the sample MAD;

- $\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} 1_{\bar{z},\infty}(z_i)$, the sample analogue of β.

Let $\widehat{\theta} = \left(\widehat{\mu}, \widehat{d}, \widehat{\beta}\right)^{\top}$ and $\theta = (\mu, d, \beta)^T$. Then we have

$$\sqrt{n}(\widehat{\theta} - \theta) = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}(z_i) + o_p(1),$$

with $\widetilde{\psi}(z) = (\widetilde{\psi}_\mu(z), \widetilde{\psi}_d(z), \widetilde{\psi}_\beta(z))^T$ defined by

$$\begin{aligned}
\widetilde{\psi}_\mu(z) &= z - \mu, \\
\widetilde{\psi}_d(z) &= 2\left((z-\mu) + ([z1_{(\mu,\infty)}(z) - z\beta] - \frac{1}{2}d) - \mu(1_{(\mu,\infty)}(z) - \beta)\right), \\
\widetilde{\psi}_\beta(z) &= (1_{(\mu,\infty)}(z) - \beta) - (z-\mu)p(\mu),
\end{aligned}$$

where $p(\mu)$ stands for the density function of z evaluated at μ, assuming that $\mathbb{P}$ represents a continuous distribution (in which case $p(\cdot)$ is assumed to be continuous in a neighborhood of μ). The expression for $\widetilde{\psi}_\mu(z)$ is standard. The expression $\widetilde{\psi}_d(z)$ is based on Gastwirth (1974). The expression $\widetilde{\psi}_\beta(z)$ follows from arguments presented in Gastwirth (1974). As a consequence, we find for the limit distribution of $\widehat{\theta}$:

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d N(0, \mathrm{cov}(\widetilde{\psi})). \tag{4.54}$$

The asymptotic covariance matrix $\mathrm{cov}(\widetilde{\psi}) = E(\widetilde{\psi}(z)\widetilde{\psi}(z)^T)$ can be estimated consistently by

$$\widehat{\mathrm{cov}}(\psi) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\widetilde{\psi}}(z_i)\widehat{\widetilde{\psi}}(z_i)^T,$$

with $\widehat{\widetilde{\psi}}(z_i)$ obtained from $\widetilde{\psi}(z_i)$ by replacing μ, d, and β by their estimates $\widehat{\mu}$, $\widehat{d}$, and $\widehat{\beta}$, and where $p(\cdot)$ is replaced by some (appropriately chosen) consistent estimator $\widehat{p}(\cdot)$.

We now proceed to the proper estimation of the parameters of the distribution of z. The parameters satisfy the bounds

$$a \le \mu \le b, \quad 0 \le d \le d_{\max}, \quad \underline{\beta} \le \beta \le \overline{\beta},$$

with

$$r_\beta = \overline{\beta} - \underline{\beta} = 4 - \frac{1}{2}\frac{d(b-a)}{(\mu-a)(b-\mu)}.$$

We can estimate $d_{\max}$ consistently by $\widehat{d}_{\max}$ (by estimating μ by $\widehat{\mu}$) and r_β consistently by $\widehat{r}_\beta$ (by estimating μ by $\widehat{\mu}$ and d by $\widehat{d}$). If $\widehat{d}_{\max}$ is not significantly different from 0, then there is not much empirical support for assuming that we 'know' d. Similarly, if $\widehat{r}_\beta$ is not significantly different from 0, then there is not much empirical support

for assuming that we 'know' β. The (asymptotic) accuracy of $\widehat{d}_{\max}$ and $\widehat{r}_\beta$ can easily be quantified using the 'delta method', resulting in $\sqrt{n}\left(\widehat{d}_{\max} - d_{max}\right) \to_d N(0, \sigma^2_{d_{\max}})$ and $\sqrt{n}\,(\widehat{r}_\beta - r_\beta) \to_d N(0, \sigma^2_{r_\beta})$.[4] With these definitions, we present now our procedure for estimation of the information basis for the use of the bounds:

1. Estimate μ by $\widehat{\mu}$, and quantify the accuracy of the latter (using the limit distribution given in (4.54)). Decide whether the accuracy is high enough to proceed under the assumption of a 'known' μ. If so, go to step 2.

2. Test the hypothesis $H_0 : d_{\max} = 0$ against $H_1 : d_{\max} > 0$, using as test statistic $\widehat{d}_{\max}/\sqrt{\widehat{\sigma}_{d_{\max}}/n}$. This is a one-sided test. If H_0 is rejected (H_1 accepted), go to step 3.

3. Estimate d by $\widehat{d}$, and quantify the accuracy of the latter (using the limit distribution given in (4.54)). Decide whether the accuracy is high enough to proceed under the assumption of a 'known' d. If so, go to step 4.

4. Test the hypothesis $H_0 : r_\beta = 0$ against $H_1 : r_\beta > 0$, using as test statistic $\widehat{r}_\beta/\sqrt{\widehat{\sigma}_{r_\beta}/n}$. This is a one-sided test. If H_0 is rejected (H_1 accepted), go to step 5.

5. Estimate β by $\widehat{\beta}$, and quantify the accuracy of the latter (using the limit distribution given in (4.54)). Decide whether the accuracy is high enough to proceed under the assumption of a 'known' β.

It may turn out that credible information is available only about the support, or support and the mean of z. In the first case, when only the support-including interval $[a, b]$ is known, a larger sample is needed to estimate other parameters. In case the support $[a, b]$ and μ are known, one may use the results of Edmundson-Madansky for the upper bound (see Remark 4.1) and Jensen for the lower bounds (see Remark 4.2).

4.B Safe approximations of chance constraints

In this Appendix we list the relevant results from Ben-Tal et al. (2009) used to prove Theorems 4.1 and 4.2, and 4.3 adopted to the notation of this paper.

[4] The 'delta method' yields $\sigma^2_{d_{max}} = r^2 \mathrm{var}(\widetilde{\psi}_\mu)$, with $r = \frac{\partial d_{\max}}{\partial \mu} = \frac{2(b+a-2\mu)}{b-a}$. Similarly, we have $\sigma^2_{r_\beta} = s^T \mathrm{cov}((\widetilde{\psi}_\mu, \widetilde{\psi}_d)^T)s$, with $s = \frac{\partial r_\beta}{\partial(\mu,d)^T} = \left(-\frac{d(b-a)(b+a-2\mu)}{2((\mu-a)(b-\mu))^2}, -\frac{b-a}{2(\mu-a)(b-\mu)}\right)^T$.

4.B.1 Safe approximation in Theorem 1

In the proof of Theorem 4.1 the following result is used.

Theorem 4.4 (Ben-Tal et al. (2009), Theorem 2.4.4) *Assume that:*

P.1. z_i, $i, \ldots, n_{\boldsymbol{z}}$ *are independent random variables such that* $supp(z_i) \subseteq [a_i^-, a_i^+]$, $i = 1, \ldots, n_{\boldsymbol{z}}$,

P.2. the distributions $\mathbb{P}_i$ *of the components* z_i *are such that*

$$\int \exp(ts) d\mathbb{P}_i(s) \leq \exp\left(\max\{\mu_i^+ t, \mu_i^- t\} + \frac{1}{2}\sigma_i^2 t\right), \quad \forall t \in \mathbb{R}, \tag{4.55}$$

with known constants $\mu_i^- \leq \mu_i^+$.

Then, the robust constraint

$$\boldsymbol{a}^T(\boldsymbol{z})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{z}), \quad \forall \boldsymbol{z} \in \mathcal{U}, \text{ where } [\boldsymbol{a}(\boldsymbol{z}); \boldsymbol{b}(\boldsymbol{z})] = [\boldsymbol{a}^0; \boldsymbol{b}^0] + \sum_{i=1}^{n_{\boldsymbol{z}}} z_i [\boldsymbol{a}_i^0; \boldsymbol{b}_i^0], \tag{4.56}$$

and

$$\mathcal{U} = \left\{ \boldsymbol{z} \in \mathbb{R}^{n_{\boldsymbol{z}}} : \exists \boldsymbol{u} \in \mathbb{R}^{n_{\boldsymbol{z}}} : \begin{array}{c} \mu_i^- \leq z_i - u_i \leq \mu_i^+, \quad i = 1, \ldots, n_{\boldsymbol{z}} \\ \sqrt{\sum_{i=1}^{n_{\boldsymbol{z}}} \frac{u_i^2}{\sigma_i^2}} \leq \sqrt{2\log(1/\epsilon)} \\ a_i^- \leq z_i \leq a_i^+, \quad i = 1, \ldots, n_{\boldsymbol{z}} \end{array} \right\},$$

is a safe approximation of (4.42). Moreover, $\boldsymbol{x}$ *satisfies (4.56) if and only if there exist* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{n_{\boldsymbol{z}}+1}$ *such that* $\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}$ *satisfy the following set of constraints::*

$$\begin{cases} (\boldsymbol{a}^i)^T \boldsymbol{x} - b^i = u_i + v_i, \quad i = 0, \ldots, n_{\boldsymbol{z}} \\ u_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} \max\left\{a_i^+ u_i, a_i^- u_i\right\} \leq 0 \\ v_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} \max\left\{\mu_i^+ v_i, \mu_i^- v_i\right\} + \sqrt{2\log(1/\epsilon)} \sqrt{\sum_{i=1}^{n_{\boldsymbol{z}}} \sigma_i^2 v_i^2} \leq 0. \end{cases}$$

4.B.2 Safe approximation in Theorem 2

The proof of Theorem 4.2 relies on the following result from Ben-Tal et al. (2009).

Theorem 4.5 (Ben-Tal et al. (2009), Proposition 4.2.2) *Assume that the distribution* $\mathbb{P}$ *of the random perturbation* $\boldsymbol{z}$ *is such that*

$$\log\left(\mathbb{E}\exp\left(\boldsymbol{w}^T\boldsymbol{z}\right)\right) \leq \Phi(\boldsymbol{w}),$$

where $\boldsymbol{w} = (w_1, \ldots, w_{n_{\boldsymbol{z}}})$ *for some known convex function* $\Phi(\cdot)$ *that is finite everywhere and satisfies* $\Phi(0) = 0$*. Then, any* $(w_0, \boldsymbol{w})$ *feasible for*

$$\inf_{\beta>0} \left\{ w_0 + \beta\Phi\left(\beta^{-}1\boldsymbol{w}\right) + \beta\log(1/\epsilon) \right\} \leq 0$$

is feasible for the chance constraint

$$\mathbb{P}\left(w_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} w_i z_i > 0 \right) \leq \epsilon.$$

Proof.[Theorem 4.2.] We show that the function $\Phi(\boldsymbol{w})$:

$$\Phi(\boldsymbol{w}) = \log(\Psi(\boldsymbol{w})), \quad \Psi(\boldsymbol{w}) = \sup_{\mathbb{P}\in\mathcal{P}_{(\mu,\boldsymbol{d})}} \mathbb{E}_{\mathbb{P}} \exp(\boldsymbol{w}^T\boldsymbol{z}) = \prod_{l=1}^{n_{\boldsymbol{z}}} \left(d_i \cosh(w_i) + 1 - d_i\right).$$

satisfies the conditions of Theorem 4.5. Indeed, from Ben-Tal and Hochman (1972) we know that $\Psi(\boldsymbol{w})$ gives a tight upper bound on $\mathbb{E}_{\mathbb{P}} \exp(\boldsymbol{w}^T\boldsymbol{z})$. Also, the function $\Phi(\boldsymbol{w})$ is convex as it is the log-sum-exp function, see Boyd and Vandenberghe (2004), and it holds that $\Phi(0) = 0$. Thus, it is sufficient to substitute

$$w_i := (\boldsymbol{a}^i)^T\boldsymbol{x} - b_i, \quad i = 0, \ldots, n_{\boldsymbol{z}},$$

to arrive at constraint (4.47) from Theorem 4.2. ■

4.B.3 Safe approximation in Theorem 3

Theorem 4.3 follows from the following result from Ben-Tal et al. (2009):

Theorem 4.6 (Ben-Tal et al. (2009), Proposition 4.3.1) *Consider a generating function* $\gamma(s)$ *satisfying (4.49). Let* $\Psi(\boldsymbol{w})$ *be a finite convex function satisfying*

$$\Psi(\boldsymbol{w}) \geq \mathbb{E}_{\mathbb{P}}\left(\gamma\left(w_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} w_i z_i \right) \right), \quad \Psi(\boldsymbol{w} + t[-1, 0, \ldots, 0]) \to 0, \text{ when } t \to \infty.$$

Then, the inequality

$$\inf_{\beta>0} \left(\beta\Psi(\beta^{-1}\boldsymbol{w}) - \beta\epsilon \right) \leq 0$$

is a safe approximation of the chance constraint

$$\mathbb{P}\left(w_0 + \sum_{i=1}^{n_{\boldsymbol{z}}} w_i z_i > 0 \right) \leq \epsilon.$$

Proof.[Theorem 4.3.] The result follows from using $\Psi(\boldsymbol{w})$ defined as in (4.50). This function clearly satisfies the conditions of Theorem 4.6. Then, the only remaining part is substituting the relevant terms for $w_i, i = 0, \ldots, n_{\boldsymbol{z}}$. ■

4.C Expectation of $\exp(\boldsymbol{w}^T\boldsymbol{z})$ without independent components

We now consider obtaining an upper bound on $\exp(\boldsymbol{w}^T\boldsymbol{z})$ using the results of Wiesemann et al. (2014), where the components of the random variable $\boldsymbol{z}$ are not assumed to be independent. For that reason, the distributional uncertainty set is given by:

$$\mathcal{P}' = \{\mathbb{P}: \quad \text{supp}(z_i) \subseteq [-1,1], \quad \mathbb{E}_{\mathbb{P}} z_i = 0, \quad \mathbb{E}_{\mathbb{P}}|z_i| = d_i, \quad i = 1,\ldots,n_{\boldsymbol{z}}\}.$$

To obtain the worst-case expectation, one needs to solve the following problem:

$$\begin{array}{ll} \min\limits_{t} & t \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}} \exp(\boldsymbol{z}^T\boldsymbol{w}) \leq t, \quad \forall \mathbb{P} \in \mathcal{P}' \end{array} \tag{4.57}$$

The uncertainty set for the distributions $\mathbb{P}$ in their framework is:

$$\mathcal{P}' = \left\{ \mathbb{P}: \begin{array}{c} \mathbb{E}_{\mathbb{P}} \left(\begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix} \boldsymbol{z} + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{u} \right) = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{d} \end{bmatrix} \\ \mathbb{P}((\boldsymbol{z},\boldsymbol{u}) \in \mathcal{C}) = 1 \end{array} \right\}, \tag{4.58}$$

where $\mathcal{C} = \{(\boldsymbol{z},\boldsymbol{u}): \quad -\boldsymbol{1} \leq \boldsymbol{z} \leq \boldsymbol{1}, \quad \boldsymbol{u} \geq \boldsymbol{z}, \quad \boldsymbol{u} \geq -\boldsymbol{z}, \quad \boldsymbol{u} \leq \boldsymbol{1}, \quad \boldsymbol{u} \geq \boldsymbol{0}\}$. Then, the problem to solve is equivalent to:

$$\begin{array}{lll} \min\limits_{\boldsymbol{\kappa},\boldsymbol{\lambda}\geq 0,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,t} & t & \\ \text{s.t.} & \boldsymbol{\beta}_2^T\boldsymbol{d} + \boldsymbol{1}^T(\boldsymbol{\kappa}-\boldsymbol{\lambda}) \leq t & \\ & \boldsymbol{z}^T\boldsymbol{\beta}_1 + \boldsymbol{u}^T\boldsymbol{\beta}_2 + \boldsymbol{1}^T(\boldsymbol{\kappa}-\boldsymbol{\lambda}) \geq \exp(\boldsymbol{z}^T\boldsymbol{w}), & \forall (\boldsymbol{z},\boldsymbol{u}) \in \mathcal{C} \end{array} \tag{4.59}$$

The last line of (4.59) involves a constraint on the function $\exp(\boldsymbol{z}^T\boldsymbol{x})$ over $\mathcal{C}$. Since $\exp(\boldsymbol{z}^T\boldsymbol{x})$ is strictly convex in $\boldsymbol{z}$, an equivalent reformulation of such a constraint would have to take into account all $3^{n_{\boldsymbol{z}}}$ vertices of $\mathcal{C}$. The number $3^{n_{\boldsymbol{z}}}$ comes from the fact that per component, the uncertainty set is a triangle $\mathcal{C}_i = \{(z_i,u_i): \quad -1 \leq z_i \leq 1, \quad u_i \geq z_i, \quad u_i \geq -z_i, \quad u_i \leq 1\}$.

4.D Properties of the MAD

In this Appendix we provide some properties of the MAD and the formulas for several well-known probability distributions, based on Ben-Tal and Hochman (1985).

If we denote by σ^2 the variance of the random variable z, whose distribution is known to belong to the set $\mathcal{P}_{(\mu,d)}$ (see 4.6), then it holds that:

$$\frac{d^2}{4\beta(1-\beta)} \leq \sigma^2 \leq \frac{d(b-a)}{2}.$$

In particular, since

$$d^2 \leq 4\beta(1-\beta)\sigma^2 \leq \sigma^2,$$

it holds that $d \leq \sigma$. For a proof, we refer the reader to Ben-Tal and Hochman (1985). For several specific distibutions, an explicit formula for d is available:

- Uniform distribution on $[a, b]$:

$$d = \frac{1}{4}(b-a)$$

- Normal distribution $N(\mu, \sigma^2)$:

$$d = \sqrt{\frac{2}{\pi}}\sigma$$

- Gamma distribution with parameters λ and k (for which $\mu = k/\lambda$):

$$d = \frac{2k^k}{\Gamma(k)\exp(k)}\frac{1}{\lambda}.$$

Ben-Tal and Hochman (1985) provide an explicit formula for the MAD for general *stable distributions*. A stable distribution is defined via its *location parameter* κ, *scale parameter* $D > 0$, *measure of skewness* $-1 \leq \lambda \leq 1$, and *characteristic exponent* $0 < \alpha \leq 2$. Important distributions belonging to this class are, for example, the normal and Cauchy distributions. The characteristic function of a stable distribution is given by

$$\log \Psi_z(t) = \log \mathbb{E}\exp(\imath t z) = \imath\kappa t - D|t|^\alpha \left(1 + \imath\lambda \text{sign}(t)\tan\left(\frac{1}{2}\pi\alpha\right)\right).$$

Stable distributions are are the only possible limiting laws for sums of independent identically distributed random variables. For properties of the stable distributions we refer the reader to Ben-Tal and Hochman (1985), who prove that for $1 < \alpha \leq 2$ the MAD of a stable random variable is given by:

$$d = D^{1/\alpha}H(\lambda, \alpha),$$

where

$$H(\lambda,\alpha) = \frac{2}{\pi}\frac{\Gamma(1-1/\alpha)}{(1+A^2)^{(\alpha-1)/2a}}\left(\cos\left((1-1/\alpha)\arctan A\right) + A\sin\left((1-1/\alpha)\arctan A\right)\right),$$

and $A = \lambda\tan\left(\frac{1}{2}\alpha\pi\right)$. In case of $\alpha \leq 1$ the mean of the random variable z does not exist.

4.E Worst-case expectations: synthesis of antenna array

In this companion we illustrate the use of the (μ, d) results in the context of incorporation of the implementation error in problems with nonlinear constraints. We consider the antenna design problem from Section 7.1.2 of Ben-Tal et al. (2009). We first introduce some necessary properties of antenna design.

The directional distribution (radiation pattern) of energy sent by a single antenna can be described in terms of an *antenna diagram* which is a complex-valued function. Its interpretation (in polar coordinates) is that the modus of the diagram stands for the amplitude of the radiation intensity at a given (fixed) distance whereas the angle of the complex number stands for the wave length (frequency). The modulus of the diagram can be changed by the amount of power allocated to the given antenna. If a device consists of more than one antenna, its diagram is a sum of the diagrams of the particular antennas. Therefore, it is possible to manipulate the power allocated to multiple antennas so that the diagram of an entire device is as close to (some) desired function as possible.

In this problem, n harmonic oscillators are placed at the points $k\boldsymbol{i}$, $k = 1, \ldots, n$, with $\boldsymbol{i}$ being the unit vector in the direction of the x-axis in $\mathbb{R}^3$. The objective is to concentrate the energy sent by the antennas within a certain region of the 3-D space, defined using the angle that 3-D directions make with the x axis. The diagram of the k-th antenna sent in direction e is given by:

$$D_k(\varphi) = \exp\left(2\pi\imath \cos(\varphi)k/\lambda\right),$$

where φ is the angle between direction e and the direction $\boldsymbol{i}$ of the X-axis, λ is the wavelength, and $\imath$ is the imaginary unit. With complex weights vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{C}^n$, the diagram of the antenna array is the sum of diagrams of the antennas:

$$D(\varphi) = \sum_{k=1}^{n} x_k D_k(\varphi).$$

Energy sent by an antenna in the direction given by an angle φ from the x-axis is proportional to the L_2 norm of the diagram. The objective is to send as much energy as possible into the region $\varphi \in [0, \Delta]$ by minimizing the weighted L_2 norm of the diagram $D(\cdot)$ in the sidelobe angle ($\mathcal{SA}$) $\Delta \leq \varphi \leq \pi$:

$$\|D(\cdot)\|_{\mathcal{SA}} = \left(\frac{1}{1+\cos(\Delta)} \int_{\Delta}^{\pi} |D(\varphi)| \sin(\varphi) d\varphi\right)^{1/2}.$$

The quantity $\|D(\cdot)\|_{\mathcal{SA}}$ can also be formulated as $\|\boldsymbol{A}\boldsymbol{x}\|$ where $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ is such that

$$\boldsymbol{A} = \boldsymbol{H}^{1/2}, \quad \boldsymbol{H} \in \mathbb{C}^{n\times n} : H_{pq} = \frac{1}{1+\cos\Delta} \int_{\Delta}^{\pi} D_p(\varphi)\overline{D_q(\varphi)} \sin(\varphi) d\varphi.$$

For the problem to be bounded, a normalization restriction is added: $\Re\left(D(0)\right) \geq 1$, where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary parts of a complex number. Weights x_k represent the electric power sent to each of the antennas and as such, are subject to implementation error. We assume that the weights x_k are distorted by the relative implementation error $\eta_k \in \mathbb{C}$ in the following fashion:

$$x_k \mapsto (1+\eta_k)x_k.$$

We assume that the real and imaginary parts of the implementation error are independent random variables with supports included in the interval $[-\rho, \rho]$, with mean 0 and MAD equal to $\theta\rho$:

$$\begin{aligned}\mathcal{P} \;=\; \{\mathbb{P} : \operatorname{supp}(\Im(\eta_k)), \operatorname{supp}(\Re(\eta_k)) \subset [-\rho,\rho], \quad & \mathbb{E}_{\mathbb{P}}\Re(\eta_k) = \mathbb{E}_{\mathbb{P}}\Re(\eta_k) = 0, \\ & \mathbb{E}_{\mathbb{P}}|\Im(\eta_k)| = \mathbb{E}_{\mathbb{P}}|\Re(\eta_k)| = \theta\rho, \quad \Im(\eta_k) \perp \Re(\eta_k), \quad k = 1,\ldots,n\}.\end{aligned}$$

The optimization problem is:

$$\begin{aligned}\min_{\tau,\boldsymbol{x}} \quad & \tau \\ \text{s.t.} \quad & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\|\boldsymbol{A}\boldsymbol{x}(\boldsymbol{\eta})\| \leq \tau \\ & \Re\left(\sum_{k=1}^{n} x_k(1+\eta_k)D_k(0)\right) \geq 1, \quad \forall \boldsymbol{\eta} \in \operatorname{supp}(\boldsymbol{\eta}),\end{aligned} \tag{4.60}$$

where $\boldsymbol{x}(\boldsymbol{\eta}) = [x_1(1+\eta_1), \ldots, x_n(1+\eta_n)]^T$, $\operatorname{supp}(\boldsymbol{\eta}) = \operatorname{supp}(\eta_1) \times \ldots \times \operatorname{supp}(\eta_n)$. The second constraint in the problem can be reformulated as a deterministic constraint:

$$\Re\left(\sum_{k=1}^{n} x_k D_k(0)\right) \geq 1 + \rho\sum_{k=1}^{n}\left|\Re\left(x_k D_k(0)\right)\right| + \rho\sum_{k=1}^{n}\left|\Im\left(x_k D_k(0)\right)\right|.$$

We solve problem (4.60) with $n = 5$ antennas, wavelength $\lambda = 8$ and $\Delta = \pi/6$ in two ways:

- **nominal:** in this case we assume $\rho = 0$ (no implementation error)
- **robust:** we assume $\rho = 0.01$ (that is, implementation error of 1%) and $\theta = 0.5$.

To compare the nominal and robust solutions, we sample uniformly 10^4 random perturbations $\hat{\boldsymbol{\eta}}$ from the set $\mathcal{E}(\widehat{\rho}) = \{\boldsymbol{\eta} : -\widehat{\rho} \leq \Re(\eta_k), \Im(\eta_k) \leq \widehat{\rho}, \quad k = 1,\ldots,n\}$, with $\widehat{\rho} \in \{0.01, 0.03, 0.05, 0.1\}$ and compute the value $\|D(\cdot)\|_{\mathcal{SA}}$ for $\boldsymbol{x}(\hat{\boldsymbol{\eta}})$. Since the normalization condition may not hold with perturbation, we normalize the diagrams $D(\cdot)$ in such a way that $|D(0)| = 1$. Table 4.10 presents the results.

The nominal solution performs well only in case of no implementation error, yielding an average value of 0.204, compared to 0.260 for the robust solution. However, already

Table 4.10 – Results of the antenna design experiment. The numbers in the columns are the mean values of simulated $\|D(\cdot)\|_{\mathcal{SA}}$ (to be minimized in the optimization problem). The numbers in brackets are standard deviations.

	Simulated $\|D(\cdot)\|_{\mathcal{SA}}$	
Solution	Nominal	Robust
$\widehat{\rho} = 0$	**0.204 (0.00)**	**0.260 (0.00)**
$\widehat{\rho} = 0.01$	**0.424 (0.19)**	**0.262 (0.00)**
$\widehat{\rho} = 0.03$	1.107 (1.41)	0.278 (0.01)
$\widehat{\rho} = 0.05$	1.223 (1.32)	0.308 (0.03)
$\widehat{\rho} = 0.1$	1.277 (1.78)	0.424 (0.13)

with the relative implementation error equal to 1%, the robust solution performs significantly better, yielding an average value 0.262 (st. dev. 0.0016), compared to 0.424 (0.19) for the nominal solution. This relationship grows even bigger for larger error values, compare 1.277 (1.78) to 0.424 (0.13) in case of 10% relative implementation error. This illustrates that the (μ, d) results provide a good way of tackling the implementation error in nonlinear constraints in a distributionally robust way.

4.F MAD of $\boldsymbol{a}^T\boldsymbol{z}$ using the results of Wiesemann et al. (2014)

Problem (4.35) is equivalent, in line with the methodology of Wiesemann et al. (2014), to:

$$\sup_{\mathbb{P}_{(\boldsymbol{z},\boldsymbol{u})} \in \mathcal{P}'} \mathbb{E}_{\mathbb{P}_{(\boldsymbol{z},\boldsymbol{u})}} \max\{\boldsymbol{a}^T\boldsymbol{z} - \boldsymbol{a}^T\boldsymbol{\mu}, -\boldsymbol{a}^T\boldsymbol{z} + \boldsymbol{a}^T\boldsymbol{\mu}\} \tag{4.61}$$

where

$$\mathcal{P}' = \left\{ \mathbb{P} : \begin{array}{c} \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u}\right) = \boldsymbol{b} \\ \mathbb{P}((\boldsymbol{z},\boldsymbol{u}) \in \mathcal{C}) = 1 \end{array} \right\}, \quad \mathcal{C} = \left\{(\boldsymbol{z},\boldsymbol{u}) : \quad \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{c}\right\},$$

where the vector $\boldsymbol{u} \in \mathbf{R}^{n_{\boldsymbol{z}}}$ consists of components u_i, each of which is an auxiliary analysis variable corresponding to the MAD of z_i. The first (moment condition) in the definition of $\mathcal{P}'$ should ensure that the first moment of $\boldsymbol{z}$ is equal to $\boldsymbol{\mu}$ and the first moment of $\boldsymbol{u}$ is equal to $\boldsymbol{d}$. We define thus:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I} \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{d} \end{bmatrix}.$$

The second (support) condition in the definition of $\mathcal{P}'$ should ensure that the support of $\boldsymbol{z}$ is the unit box and that $\boldsymbol{u}$ indeed corresponds to the deviation of $\boldsymbol{u}$. We need

to ensure thus that:

$$\|\boldsymbol{z}\|_\infty \leq 1, \quad \boldsymbol{u} \geq \boldsymbol{z} - \boldsymbol{\mu}, \quad \boldsymbol{u} \geq \boldsymbol{\mu} - \boldsymbol{z}, \quad \boldsymbol{u} \geq 0, \quad \boldsymbol{u} \leq 1,$$

where the last condition ensures boundedness of $\mathcal{C}$, required by Wiesemann et al. (2014). We ensure these conditions by setting:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{I} \\ -\boldsymbol{I} \\ \boldsymbol{I} \\ -\boldsymbol{I} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ -\boldsymbol{I} \\ -\boldsymbol{I} \\ \boldsymbol{I} \\ -\boldsymbol{I} \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{1} \\ \boldsymbol{\mu} \\ -\boldsymbol{\mu} \\ \boldsymbol{1} \\ \boldsymbol{0} \end{bmatrix}.$$

Wiesemann et al. (2014) prove (Theorem 1 in their paper) that under mild conditions satisfied in our case, (4.61) is equivalent to the following LP:

$$\begin{aligned} \min_{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \geq 0, w, \boldsymbol{\beta}, \kappa} \quad & w \\ \text{s.t.} \quad & \boldsymbol{b}^T \boldsymbol{\beta} + \kappa \leq w \\ & \boldsymbol{c}^T \boldsymbol{\phi}_1 - \boldsymbol{a}^T \boldsymbol{\mu} \leq \kappa \\ & \boldsymbol{c}^T \boldsymbol{\phi}_2 + \boldsymbol{a}^T \boldsymbol{\mu} \leq \kappa \\ & \boldsymbol{C}^T \boldsymbol{\phi}_1 + \boldsymbol{A}^T \boldsymbol{\beta} = \boldsymbol{a} \\ & \boldsymbol{C}^T \boldsymbol{\phi}_2 + \boldsymbol{A}^T \boldsymbol{\beta} = -\boldsymbol{a} \\ & \boldsymbol{D}^T \boldsymbol{\phi}_1 + \boldsymbol{B}^T \boldsymbol{\beta} = 0 \\ & \boldsymbol{D}^T \boldsymbol{\phi}_2 + \boldsymbol{B}^T \boldsymbol{\beta} = 0. \end{aligned} \tag{4.62}$$

Bibliography

R. T. Baillie, G. G. Booth, Y. Tse, and T. Zabotina. Price discovery and common factor models. *Journal of Financial Markets*, 5(3):309–321, 2002.

A. Ben-Tal and D. den Hertog. Immunizing conic quadratic optimization problems against implementation errors. *Mathematical Programming*, 149(1-2):265–299, 2011.

A. Ben-Tal and E. Hochman. More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability*, 9:803–812, 1972.

A. Ben-Tal and E. Hochman. Approximation of expected returns and optimal decisions under uncertainty using mean and mean absolute deviation. *Zeitschrift für Operations Research*, 29(7):285–300, 1985.

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

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

A. Ben-Tal, B. Golany, A. Nemirovski, and J.-Ph. 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 University Press, 2009.

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

A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.

D. Bertsimas, V. Goyal, and X. A. Sun. A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. *Mathematics of Operations Research*, 36(1):24–54, 2011.

J. R. Birge and R. J.-B. Wets. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research*, 12(1): 149–162, 1987.

S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2007.

F. De Ruiter, D. den Hertog, and R. Brekelmans. The impact of the existence of multiple adjustable robust solutions. *To appear in Mathematical Programming*, 2016.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

J. H. Dulá. An upper bound on the expectation of simplicial functions of multivariate random variables. *Mathematical Programming*, 55(1-3):69–80, 1992.

J. Dupačová. On minimax solutions of stochastic linear programming problems. *Časopis Pro Pětováni Matematiky*, 91:423–430, 1966.

J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95: 493–511, 2003.

N. C. P. Edirisinghe. *Stochastic Programming: The State of the Art In Honor of George B. Dantzig*, chapter Stochastic Programming Approximations Using Limited Moment Information, with Application to Asset Allocation, pages 97–138. Springer New York, 2011.

N. C. P. Edirisinghe and W. T. Ziemba. Tight bounds for stochastic convex programs. *Operations Research*, 40:660–677, 1992.

N. C. P. Edirisinghe and W. T. Ziemba. Bounds for two-stage stochastic programs with fixed recourse. *Mathematics of Operations Research*, 19:292–313, 1994a.

N. C. P. Edirisinghe and W. T. Ziemba. Bounding the expectation of a saddle function with application to stochastic programming. *Mathematics of Operations Research*, 19: 314–340, 1994b.

H. P. Edmundson. Bounds on the expectation of a convex function of a random variable. *Acta Mathematica*, 30:175–193, 1956.

E. A. H. El Amir. On uses of mean absolute deviation: decomposition, skewness and correlation coefficients. *Metron*, 70(2-3):145–164, 2012.

P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein Metric: Performance guarantees and tractable reformulations. http://www.optimization-online.org/DB_FILE/2015/05/4899.pdf, 2015.

K. Frauendorfer. SLP recourse problems with arbitrary multivariate distributions - the dependent case. *Mathematics of Operations Research*, 13:377–394, 1988.

G. Gallego. A minmax distribution free procedure for the (q, r) inventory model. *Operations Research Letters*, 11(1):55–60, 1992.

G. Gallego, J. K. Ryan, and D. Simchi-Levi. Minimax analysis for finite-horizon inventory models. *IIE Transactions*, 33(10):861–874, 2001.

H. I. Gassmann and W. T. Ziemba. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Mathematical Programming Study*, 27: 39–53, 1986.

J. L. Gastwirth. Large sample theory of some measures of income inequality. *Econometrica: Journal of the Econometric Society*, 42(1):191–196, 1974.

K. Ghoudi, R. J. Kulperger, and B. Rémillard. A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, 79(2):191–218, 2001.

J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4):902–917, 2010.

G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62, 2015.

J. Higle and S. Sen. *Stochastic decomposition: A statistical method for large scale stochastic linear programming*. Kluwer, Dordrecht, 1996.

D. A. Iancu and N. Trichakis. Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147, 2013.

J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.

R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158(1):291–327, 2016.

I. Jolliffe. *Principal component analysis*. John Wiley & Sons, Ltd., 2002.

P. Kall. An upper bound for SLP using first and total second moments. *Annals of Operations Research*, 30:267–276, 1991.

A. F. Karr. Extreme points of certain sets of probability measures, with applications. *Mathematics of Operations Research*, 8:74–85, 1983.

A. Madansky. Bounds on the expectation of a convex function of a multivariate random variable. *The Annals of Mathematical Statistics*, 30:743–746, 1959.

G. Perakis and G. Roels. Regret in the newsvendor model with partial information. *Operations Research*, 56(1):188–203, 2008.

J. Pinkse. A consistent nonparametric test for serial independence. *Journal of Econometrics*, 84(2):205–231, 1998.

I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

K. Postek, D. den Hertog, and B. Melenberg. Computationally tractable counterparts of distributionally robust constraints on risk measures. *To appear in SIAM Review.*, 2015.

W. W. Rogosinsky. Moments of non-negative mass. *Proceedings of the Royal Society of London, Series A*, 245:1–27, 1958.

H. E. Scarf. A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, and H. E. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, 1958.

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

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.

M. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.

R. L. Smith. Efficient monte-carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32:1296–1308, 1984.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

CHAPTER 5

Efficient methods for several classes of ambiguous stochastic programming problems under mean-MAD information

5.1 Introduction

In practice, many decision makers are faced with uncertainty in some parameters of their model. Consider, for example, customer demand in production planning, supply of renewable energy in unit commitment problems, the precision of physical devices in engineering design, and the return on investment in finance. In these problems, the information on the uncertain parameter, based on e.g. historical data or expert opinions, can either be limited or extensive. In the latter case, uncertainty is typically modelled using random parameters with known probability distributions, whereas in the former case the distributions are *ambiguous*, i.e. only partly known (Knight 1921). In this paper, we address mathematical optimization problems where (some of) the parameter distributions are ambiguous. One of our main contributions is that for such problems we provide efficient solution methods that are easy to implement using off-the-shelf software.

If the uncertain random parameters are revealed gradually over time, then we can model the decision problem as a multi-stage stochastic programming (SP) problem (see, e.g., the textbooks Birge and Louveaux (1997), Prékopa (1995), Shapiro et al. (2009)) in which the planning horizon consists of multiple time stages. Under the assumption that the probability distributions of the uncertain random parameters are known, the problem is to determine so-called here-and-now decisions implemented before (some of) these uncertain parameters are revealed and new decisions have to be made. This process repeats itself over several stages and the objective is to minimize the sum of the here-and-now costs and the *expected* future costs, taking the distributions of the uncertain random parameters and the decisions in later stages into account.

We, however, assume in line with practical experience that only limited information

on the distributions of the uncertain random parameters is available, and instead of minimizing the expected costs, we take a *distributionally robust* (ambiguous) approach and minimize the *worst-case expected costs* over all possible (or admissible) probability distributions.

Early contributions to the minimization of worst-case expectations are the works of Scarf (1958) and Žáčková (1966). In the SP literature it has been referred to as *minimax* problems, and is mainly considered within the framework of generalized moment problems (Kemperman 1968), where the objective is to determine the worst-case expectation of a given function under conditions on some (generalized) moments of the uncertain random parameters. For a discussion on cases in which exact bounds or approximation procedures are available, we refer to Edirisinghe (2011).

Moreover, Shapiro and Kleywegt (2002) and Shapiro and Ahmed (2004) use duality to show that distributionally robust SP problems can be equivalently reformulated as a standard SP problem in which the probability distributions of the random parameters are known. The difference with our approach is that we can utilize the explicitly known worst-case distributions under the given distributional information, instead of using duality.

Recently, distributional uncertainty has gained the attention of the Robust Optimization (RO) community. They treat the sets of admissible probability distributions as uncertainty sets and use conic duality (see the references above and also, e.g., Isii (1963), Shapiro (2001), and Ben-Tal et al. (2015)) to derive equivalent, computationally tractable forms of constraints on the worst-case expectation. Prominent examples of this approach are the papers of Delage and Ye (2010) and Wiesemann et al. (2014), who provide also good surveys of the existing approaches.

Despite these developments, solving distributionally robust SP problems remains challenging for several reasons. First, it may be hard to determine the worst-case probability distribution (maximizing the expected costs), even for given here-and-now decisions. Second, it may be computationally intractable to determine optimal here-and-now decisions taking into account all decisions in future time stages under all possible realizations of the uncertain parameter. Third, the problem may contain integer decision variables, and fourth, solving the problem may require special purpose algorithms that are not available in standard software packages.

In this paper we overcome these four challenges for a large class of distributionally robust SP problems. In the remainder of this introduction we discuss each of these major challenges separately.

In our setting, the first challenge (evaluating the worst-case expectation) is void because of the particular distributional information we use — the supports, means, mean absolute deviations from the means (MADs), and the probability that a given

variable is at least equal to its mean, which are easy to estimate using e.g. the procedures given in in Postek et al. (2015). Under such distributional information, we can use a result of Ben-Tal and Hochman (1972), to prove that the worst-case and best-case marginal distributions are discrete with at most three possible realizations if the distributionally robust SP problem only contains continuous decision variables. The well-known Edmundson-Madansky upper bound (Edmundson 1956) and Jensen lower bound (Jensen 1906) are similar in spirit to the results we use. The main difference between our results and the work of e.g. Shapiro and Kleywegt (2002) and Shapiro and Ahmed (2004) is that in our case we obtain simple worst-case and best-case distributions that are the same for every here-and-now decision. This indeed simplifies matters considerably: the problem becomes much more tractable and from the practical perspective it is much more intuitive that the distributions do not depend on the (initial) decisions. Our approach is applicable only if the random parameters are stochastically independent. This is obviously a restrictive assumption but an advantage of our approach is that it can exploit this property.

The difference between the worst-case and best-case expectations of the objective gives an easy-to-calculate upper bound on the *value of distributional information* (VDI); see e.g. Delage et al. (2015). The VDI is related to the value of the stochastic solution (VSS) introduced by Birge (1982). However, the VSS measures the added value of solving the stochastic problem instead of its deterministic version, whereas the VDI measures the added value of (or the willingness to pay for) knowing the probability distributions of the random parameters. The VDI is particularly relevant in a data-driven environment where it can be used to assess the costs of gathering more data.

If all worst-case distributions are discrete with three possible realizations (as we will find), then the distributionally robust SP problem reduces to a standard SP problem with 3^n realizations (or scenarios) of the joint distribution of the n random parameters. This exponential number of scenarios explains the second major challenge (computational tractability): it is the reason why the problem is computationally intractable from an RO point of view. For this reason, the problem is often approximated by imposing *decision rules* on future decisions as a function of the revealed random parameters; see Garstka and Wets (1974) for the first contribution in the SP literature. In the RO literature Ben-Tal et al. (2004) first formulated the decision rules as *affine functions* of the revealed parameters, and their approach has been extended to other function classes by e.g. Chen and Zhang (2009), Ben-Tal et al. (2009) and Bertsimas et al. (2011). From an SP perspective, however, dealing with 3^n scenarios is not unusual and there exist many solution methods to (approximately) solve such problems. In Section 5.4 we give a brief overview of these methods. In particular, in Appendix 5.B we present a particularly efficient implementation of such methods tailored to the problem at hand.

The third major challenge (inclusion of integer variables) is relevant since many decision problems require integer decision variables to be modelled realistically. Consider e.g. unit commitment decisions in electric power generation (see e.g. Römisch and Schultz (1996), Bertsimas et al. (2013) and many others) or lot sizing decisions in inventory control (see, e.g., Postek and den Hertog (2016)). Within the SP literature, stochastic mixed-integer programming (SMIP) problems have been studied by e.g. Laporte and Louveaux (1993), Carøe and Schultz (1999), and Ahmed et al. (2004), (see also the surveys by Schultz (2003), Klein Haneveld and van der Vlerk (1999), and Sen (2005)), while in the RO literature systematic approaches have been developed to allow for integer decision variables in future time stages; see e.g. Bertsimas and Georghiou (2015), Hanasusanto et al. (2015), and Postek and den Hertog (2016).

For SMIP problems the main difficulty is that due to the integer variables in future time stages, the optimal objective value is generally not convex in the uncertain parameter. For this reason, van der Vlerk (2004), Klein Haneveld et al. (2006), Romeijnders et al. (2015), and Romeijnders et al. (2016b) have proposed convex approximations for several classes of SMIP problems. For these approximations error bounds have been derived that depend on the total variations of the probability density functions of the random parameters in the model. We use the idea of convex approximations to provide a framework for solving a large class of two-stage distributionally robust SMIP problems in which the distributions of some random parameters are known and others are ambiguous. We derive error bounds for two approximations of which one is obtained by (incorrectly) assuming that the worst-case distributions are the same as in the continuous distributionally robust SP case, i.e., assuming convexity. In Section 5.5 we apply this framework to an operating room scheduling problem.

In that section we carry out numerical experiments on an inventory control problem as well. Dealing with the fourth major challenge (ease of implementation), we show that we can obtain good solutions using off-the-shelf software, despite the exponential number of scenarios. Moreover, we show that for problems of realistic size we may obtain exact optimal solutions, using the specific structure of the problem to speed up existing algorithms. Furthermore, we provide additional managerial insights (i) by calculating the VDI, (ii) by graphically depicting the so-called *Pareto-stripe*, an extension of the Pareto curve, which shows the tradeoff between various types of objectives, and (iii) by comparing various approaches from the SP and RO literature.

To summarize, we provide a framework for solving distributionally robust SP problems, satisfying the following properties:

1. the required parameters of the independent probability distributions in the ambiguity set can be estimated from data;

2. there is a simple worst-case distribution that is the same for all here-and-now decisions;

3. future decisions depend on the observed value of the uncertain parameter;

4. the solution method is able to accommodate for integer decision variables in two-stage problems;

5. the value of distributional information can be quantified;

6. the solution method is easy to implement using off-the-shelf software and known SP techniques.

The structure of our paper is as follows. In Section 2.1 we introduce our approach for two-stage continuous problems and we extend it to the multi-stage setting in Section 2.2. Section 5.3 includes our new theoretical results on convex approximations of two-stage stochastic programs with integer decision variables. In Section 5.4 we discuss general techniques helpful in dealing with the number of scenarios which grows exponentially with the dimension of the vector of uncertain parameters. In Section 5.5 we present three numerical experiments involving operating room planning and inventory management. Each of the experiments illustrates our approach for a particular class of distributionally robust SP problems: two-stage problems problems with continuous and with integer variables, and a continuous multi-stage problem.

5.2 Distributionally robust SP problems

In this section we describe our approach for solving distributionally robust SP problems in case all decision variables are continuous. For ease of exposition, we first consider two-stage problems in Section 5.2.1; multi-stage problems are discussed in Section 5.2.2. Although the results in this section appear to be known in the SP literature (Ben-Tal and Hochman 1976), we are the first — to our knowledge — to make these results explicit in a multi-stage setting.

5.2.1 Two-stage problems

The distributionally robust SP problem that we consider is

$$\inf_{x \in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + v(x, z)], \tag{5.1}$$

where $X = \{x \in \mathbb{R}^{n_1}_+ : Ax = b\}$ represents the set of feasible first-stage solutions, $\mathcal{P}_z$ is the ambiguity set for probability distributions, and $v(x, z)$ is the second-stage

value function defined as a function of the first-stage variables x and the random parameters $z = (\xi, \omega)$:

$$v(x, z) = \inf_{y \in Y} \Big\{ q(\xi)^\top y : Wy = h(\omega) - T(\omega)x \Big\}.$$

Here, y are the second-stage (or recourse) variables and $Y \subset \mathbb{R}^{n_2}_+$ is a polyhedral set. The second-stage costs $q(\xi)$, the technology matrix $T(\omega)$, and the right-hand side $h(\omega)$ depend on the random vector $z = (\xi, \omega)$. We assume that q, T, and h are affine functions of z and that *all components of z are independent*. Thus, in particular, $q(\xi)$ is independent from $T(\omega)$ and $h(\omega)$. Moreover, since the recourse matrix W is deterministic, we say that the problem has *fixed recourse* (see, e.g., Shapiro et al. (2009)).

In problem (5.1), the here-and-now decisions x have to be made while the parameter z is unknown, and after the uncertain parameter z is revealed we are allowed to take recourse actions y to compensate for possible violations of the constraints $T(\omega)x = h(\omega)$. The objective is to minimize the sum of the direct costs $c^\top x$ and the worst-case expected costs $\sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[v(x, z)]$.

Here, the ambiguity set $\mathcal{P}_z$ is defined as

$$\mathcal{P}_z = \Big\{ \mathbb{P}_z : \operatorname{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_{\mathbb{P}_z}[z_i] = \mu_i, \quad \mathbb{E}_{\mathbb{P}_z}|z_i - \mu_i| = d_i, \quad \mathbb{P}_z\{z_i \geq \mu_i\} = \beta_i, \\ z_i \perp z_j,\ i \neq j \Big\}, \tag{5.2}$$

where $z_i \perp z_j$ means that z_i and z_j are stochastically independent. Postek et al. (2015) explain procedures to estimate these parameters from historical data and the conditions on a, b, μ, d, β such that $\mathcal{P}_z$ is non-empty. Throughout this paper we refer to the ambiguity set $\mathcal{P}_z$ in (5.2) as a (μ, d, β) ambiguity set.

Example 5.1 *Consider an inventory manager who needs to order a specific amount x of products. Later, when the uncertain customer demand z is known, he can order an additional amount y of the products, however, at unknown but likely higher prices. The objective of the manager is to minimize the expected total cost. However, due to lack of knowledge on the true distribution, he chooses the 'safe option' and minimizes the worst-case expected cost over the set $\mathcal{P}_z$ of distributions that can be 'true' based on the data.*

5.2.1.1 Worst-case expectation

Problem (5.1) is difficult to solve because the worst-case probability distribution $\mathbb{P}_z \in \mathcal{P}_z$ may depend on the first-stage decision $x \in X$, and we need to optimize over

x. However, for the (μ, d, β) ambiguity set $\mathcal{P}_z$ in (5.2), the worst-case distribution $\mathbb{P}_{\bar{z}}$ turns out to be the same for every first-stage decision so that the distributionally robust SP problem in (5.1) reduces to

$$\inf_{x\in X} \mathbb{E}_{\mathbb{P}_{\bar{z}}}[c^\top x + v(x, \bar{z})],$$

where each component of $\bar{z}$ follows a known discrete distribution with at most three realizations. This result is summarized in Proposition 5.1 below. Its proof combines the fact that the second-stage value function $v(x, z)$ is convex in ω and concave in ξ (see, e.g., Fiacco and Kyparisis 1986) with results from Ben-Tal and Hochman (1972), who provide closed-form expressions for the worst-case expectations maximizing and minimizing the expectations of convex and concave functions.

Proposition 5.1 *The two-stage distributionally robust SP problem*

$$\inf_{x\in X} \sup_{\mathbb{P}_z\in\mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}\left[c^\top x + \inf_{y\in Y}\left\{q(\xi)^\top y : Wy = h(\omega) - T(\omega)x\right\}\right]$$

with (μ, d, β) ambiguity set $\mathcal{P}_z$ for $z = (\xi, \omega) \in \mathbb{R}^{n_\xi} \times \mathbb{R}^{n_\omega}$ as defined in (5.2) is equivalent to

$$\inf_{x\in X} \mathbb{E}_{\mathbb{P}_{\bar{z}}}\left[c^\top x + \inf_{y\in Y}\left\{q(\bar{\xi})^\top y : Wy = h(\bar{\omega}) - T(\bar{\omega})x\right\}\right],$$

where the worst-case random vector $\bar{z} = (\bar{\xi}, \bar{\omega}) \in \mathbb{R}^{n_\xi} \times \mathbb{R}^{n_\omega}$ has independent components with marginal distributions

$$\mathbb{P}\left\{\bar{\xi}_i = \mu_i - \frac{d_i}{2(1-\beta_i)}\right\} = 1 - \beta_i, \quad \textit{and} \quad \mathbb{P}\left\{\bar{\xi}_i = \mu_i + \frac{d_i}{2\beta_i}\right\} = \beta_i,\ i = 1, \ldots, n_\xi,$$

and

$$\mathbb{P}\left\{\bar{\omega}_i = a_{n_\xi+i}\right\} = \frac{d_{n_\xi+i}}{2(\mu_{n_\xi+i} - a_{n_\xi+i})}, \quad \mathbb{P}\left\{\bar{\omega}_i = b_{n_\xi+i}\right\} = \frac{d_{n_\xi+i}}{2(b_{n_\xi+i} - \mu_{n_\xi+i})},$$
$$\mathbb{P}\left\{\bar{\omega}_i = \mu_{n_\xi+i}\right\} = 1 - \frac{d_{n_\xi+i}}{2(\mu_{n_\xi+i} - a_{n_\xi+i})} - \frac{d_{n_\xi+i}}{2(b_{n_\xi+i} - \mu_{n_\xi+i})}$$

for $i = 1, \ldots, n_\omega$.

Proof. See Appendix 5.A. ■

Since the worst-case distribution $\mathbb{P}_{\bar{z}}$ has finite support, we can enumerate all $K = 2^{n_\xi} \times 3^{n_\omega}$ scenarios of $\bar{z}$ and rewrite the distributionally robust SP problem in (5.1) as

$$\begin{aligned}\inf_{x\in X} \sup_{\mathbb{P}_z\in\mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + v(x, z)] &= \inf_{x\in X} \mathbb{E}_{\mathbb{P}_{\bar{z}}}[c^\top x + v(x, \bar{z})] \\ &= \inf_{x\in X} \sum_{k=1}^{K} p_k[c^\top x + v(x, \bar{z}^k)],\end{aligned}$$

where p_k denotes the probability of scenario $\bar{z}^k$, $k = 1, \ldots, K$. The latter problem can be rewritten in its deterministic equivalent form, yielding

$$\inf_{x \in X, y_k \in Y} \left\{ c^\top x + \sum_{k=1}^{K} p_k q(\bar{\xi}^k) y_k : W y_k = h(\bar{\omega}^k) - T(\bar{\omega}^k) x,\ k = 1, \ldots, K \right\}.$$

Remark 5.1 *Note that the worst-case expectation of the random vector ω does not require information on parameter β, i.e. in case one deals with uncertainty in the constraints of the problems only, it suffices to estimate the parameter a, b, μ, and d of the probability distribution ω. This means that having the knowledge on β does not change the worst-case expectation value. An even more striking fact is that if β is known, the three point worst-case distribution of $\boldsymbol{\omega}$ may not satisfy this probability bound, i.e., it may hold that $\mathbb{P}(\bar{\omega}_i \geq \mu_{n_\xi + i}) < \beta_{n_\xi + i}$ for some $1 \leq i \leq n_\omega$. This is because the worst-case probability bound is tight but it need not be attained, see Ben-Tal and Hochman (1972).*

5.2.1.2 Best-case expectation

Similar as for the worst-case expectation we can obtain the best-case expectation over all probability distributions in the (μ, d, β) ambiguity set $\mathcal{P}_z$ by using results of Ben-Tal and Hochman (1972). Again, the best-case distribution $P_{\underline{z}}$ is a discrete distribution with at most three realizations per component that does not depend on the first-stage decision x.

Proposition 5.2 *The two-stage distributionally robust SP problem*

$$\inf_{x \in X} \inf_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z} \left[c^\top x + \inf_{y \in Y} \left\{ q(\xi)^\top y : W y = h(\omega) - T(\omega) x \right\} \right]$$

with (μ, d, β) ambiguity set $\mathcal{P}_z$ for $z = (\xi, \omega) \in \mathbb{R}^{n_\xi} \times \mathbb{R}^{n_\omega}$ as defined in (5.2) is equivalent to

$$\inf_{x \in X} \mathbb{E}_{\mathbb{P}_{\underline{z}}} \left[c^\top x + \inf_{y \in Y} \left\{ q(\underline{\xi})^\top y : W y = h(\underline{\omega}) - T(\underline{\omega}) x \right\} \right], \tag{5.3}$$

where the best-case random vector $\underline{z} = (\underline{\xi}, \underline{\omega}) \in \mathbb{R}^{n_\xi} \times \mathbb{R}^{n_\omega}$ has independent components with marginal distributions

$$\mathbb{P}\left\{ \underline{\xi}_i = a_i \right\} = \frac{d_i}{2(\mu_i - a_{n_\xi + i})}, \qquad \mathbb{P}\left\{ \underline{\xi}_i = b_i \right\} = \frac{d_i}{2(b_i - \mu_i)},$$
$$\mathbb{P}\left\{ \underline{\xi}_i = \mu_i \right\} = 1 - \frac{d_i}{2(\mu_i - a_i)} - \frac{d_i}{2(b_i - \mu_i)}$$

for $i = 1, \ldots, n_\xi$ *and*

$$\mathbb{P}\left\{\underline{\omega}_i = \mu_{n_\xi+i} - \frac{d_{n_\xi+i}}{2(1-\beta_{n_\xi+i})}\right\} = 1 - \beta_{n_\xi+i}, \quad \mathbb{P}\left\{\underline{\omega}_i = \mu_{n_\xi+i} + \frac{d_{n_\xi+i}}{2\beta_{n_\xi+i}}\right\} = \beta_{n_\xi+i}.$$

for $i = 1, \ldots, n_\omega$.

Proof. See Appendix 5.A. ■

Notice that since $v(x, z)$ is concave in ξ and convex in ω the worst-case distribution of ξ has the same structure as the best-case distribution of ω and vice versa. Moreover, we can derive the deterministic equivalent form of (5.3) analogous to that of the worst-case expectation.

Best-case expectation is a useful complement to the worst-case expectation since they bound the actual expected costs in the stochastic program, which is unknown since the probability distribution of the random vector of parameters z is unknown. The difference between the worst-case and best-case expectation can be interpreted as an upper bound on the *value of distributional information* (VDI, Delage et al. (2015)), i.e. the amount we are willing to pay for complete knowledge of the probability distribution of z. We illustrate this concept in the numerical experiments of Section 5.5.

5.2.2 Multi-stage problems

We consider now the general multi-stage linear problem. For ease of exposition we limit ourselves to the uncertainty in the constraints driven by random vector $z = \omega$. The results, however, extend easily to the case including also uncertainty in the cost vector driven by a random vector ξ, as in the two-stage problem (5.1).

Here, $x_t \in \mathbb{R}^{n_t}_+$ denote the decision vectors implemented at time $t = 1, 2, \ldots, T$. The uncertain parameter $z \in \mathbb{R}^{n_z}$ has a corresponding structure $z = (z_1, \ldots, z_{T-1})$, $z_t \in \mathbb{R}^{n_{z,t}}$ for $t = 1, \ldots, T-1$, with $n_z = \sum_{t=1}^{T-1} n_{z,t}$. The time sequence of decisions and uncertainty revealing is

$$x_1 \quad \to \quad z_1 \quad \to \quad x_2 \quad \to \quad z_2 \quad \to \quad \ldots \quad \to \quad z_{T-1} \quad \to \quad x_T.$$

Since all random parameters are independently distributed, our formulation of the multi-stage case has a nested form as in Shapiro et al. (2009). The problem to solve at time $t = 1$ is:

$$\inf_{x_1 \in X_1} \left\{ c_1^\top x_1 + \sup_{\mathbb{P}_{z_1} \in \mathcal{P}_{z_1}} \mathbb{E}_{\mathbb{P}_{z_1}} v_2(x_1, z_1) \right\}, \tag{5.4}$$

where $X_1 = \{x_1 \in \mathbb{R}^{n_1}_+ : \ A_{11}x_1 = b_1\}$ and the ambiguity set $\mathcal{P}_{z_t}$, $t = 1, \ldots, T-1$, is defined as

$$\mathcal{P}_{z_t} = \{\mathbb{P}_{z_t} : \quad \text{supp}(z_{ti}) \subseteq [a_{ti}, b_{ti}], \quad \mathbb{E}_{\mathbb{P}}(z_{ti}) = \mu_{ti}, \quad \mathbb{E}_{\mathbb{P}}\left|z_{ti} - \mu_{ti}\right| = d_{ti}, \quad \forall i, \\ z_{ti} \perp z_{t,j}, \quad \forall i \neq j, \}\,.$$

The value function $v_t(x_{t-1}, z_{t-1})$, $t = 2, \ldots, T-1$ is defined as the optimal value of the optimization problem to solve at time t:

$$v_t(x_{t-1}, z_{t-1}) = \inf_{x_t \in X_t} \left\{ c_t^\top x_t + \sup_{\mathbb{P}_{z_t} \in \mathcal{P}_{z_t}} \mathbb{E}_{\mathbb{P}_{z_t}} v_{t+1}(x_t, z_t) \right\}, \tag{5.5}$$

where $X_t = \{x_t \in \mathbb{R}^{n_t}_+ : \ \sum_{s=1}^{t-1} A_{ts}(z_{t-1})x_s + A_{tt}x_t = b_t(z_{t-1})\}$ and $v_T(x_{T-1}, z_{T-1})$ is the optimal value of the optimization problem at stage T:

$$v_T(x_{T-1}, z_{T-1}) = \inf_{x_T \in X_T} \left\{ c_T^\top x_T \right\}, \tag{5.6}$$

with $X_T = \{x_T \in \mathbb{R}^{n_T}_+ : \ \sum_{s=1}^{T-1} A_{Ts}(z_{T-1})x_s + A_{TT}x_T = b_T(z_{T-1})\}$. At each stage t the objective function consists of a linear component involving the decisions x_t and (with the exception of stage T) the worst-case expected value of the optimal value of the problem to be solved at the next stage. At each stage, a system of constraints is to hold that involves the decision vectors $x_1, \ldots, x_t$, and the coefficients $A_{ts}(\cdot)$, and $b_t(\cdot)$ which depend on the outcome of the uncertain parameter z_{t-1}, observed before x_t is implemented. We assume that the functions $A_{ti}(\cdot)$, and $b_t(\cdot)$ are linear. The assumption that matrices A_{tt}, $t = 1, \ldots, T$ are fixed is the multi-stage equivalent of the two-stage *fixed recourse* restriction.

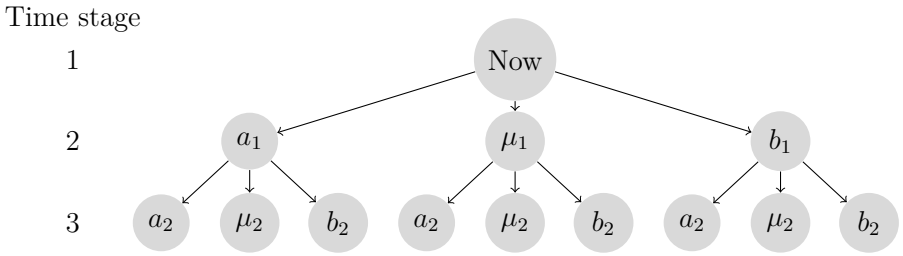
In the two-stage case of Section 5.2.1, in order to reformulate problem (5.1) to the closed-form equivalent, the function $v(x, \cdot)$ has to be convex. A similar property is needed here to reformulate the multi-stage problem to a closed form and holds for the functions $v_t(x_{t-1}, \cdot)$; moreover, at each time t the decision maker is solving a convex problem in the decision variables, as stated by the following proposition, leading to a tractable convex optimization problem.

Proposition 5.3 *Functions $v_t(x_{t-1}, z_{t-1})$ and $v_T(x_{T-1}, z_{T-1})$ are convex in z_{t-1}, $t = 2, \ldots, T-1$ and z_{T-1}, respectively, and the optimization problems (5.5) and (5.6) are convex in x_t for $t = 1, \ldots, T-1$ and in x_T, respectively.*

Proof. See Appendix 5.A. ■

Proposition 5.3 implies that we can use the results of Ben-Tal and Hochman (1972) to give a closed form of the multi-stage problem (5.4). We do this by recursively inserting the worst-case distributions of z_t from Proposition 5.1, considering the problem at stages $t = T-1, T-2, \ldots, 2$. The final result is stated in the following proposition.

Figure 5.1 – Scenario tree for the worst-case distribution of $\bar{z} \in \mathbb{R}^2$ in the multi-stage setting $\bar{z} = (\bar{z}_1, \bar{z}_2) \in \mathbb{R}^2$.



Proposition 5.4 *Distributionally robust SP problem formulated in (5.4), (5.5), and (5.6) is equivalent to the following problem:*

$$\inf_{x_1 \in X_1} \left\{ c_1^\top x_1 + \mathbb{E}_{\mathbb{P}_{\bar{z}_1}} v_2(x_1, \bar{z}_1) \right\}, \tag{5.7}$$

where

$$v_t(x_{t-1}, z_{t-1}) = \inf_{x_t \in X_t} \left\{ c_t^\top x_t + \mathbb{E}_{\mathbb{P}_{\bar{z}_t}} v_{t+1}(x_t, \bar{z}_t) \right\}, \ t = 1, \ldots, t-1, \tag{5.8}$$

where the worst-case distributions $\mathbb{P}_{\bar{z}_t}$ are defined as in Proposition 5.1, and $v_T(x_{T-1}, z_{T-1})$ is given by (5.6).

Formulations (5.7) and (5.8), together with the final-stage problem (5.6) constitute a single big optimization problem with a tree structure. In this structure, the first-stage problem refers via $v_2(x_1, \bar{z}^k)$, $k = 1, \ldots, 3^{n_{z,1}}$ to $3^{n_{z,1}}$ second-stage problems, each of which links to $3^{n_{z,2}}$ stage 3 problems, and so on. This corresponds to the tree structure of the worst-case distribution of the uncertain parameter, depicted in Figure 5.1.

Remark 5.2 *As mentioned in the beginning of this section, it is possible, similar as in the two-stage case, (i) to consider also uncertainty in the objective function coefficients $c_2, \ldots, c_T$ since the solutions of the optimization problems at each stage are concave in $c_2, \ldots, c_T$, respectively; (ii) to construct a closed form of the problem in which the best-case expectation is minimized with respect to the distribution of parameter $z = (\xi, \omega)$.*

5.3 Two-stage mixed-integer recourse models

Mixed-integer recourse models arise when the optimization problem involves integer decision variables. The advantage of incorporating such variables in the model is that they may be used to model e.g. indivisibilities or on/off decisions, the disadvantage

however is that solving the model becomes much more complicated because generally the second-stage value function is non-convex. In the distributionally robust context of this paper, this implies that the result of Ben-Tal and Hochman (1972) cannot be applied directly. Nevertheless, their result may be of use when we consider two-stage mixed-integer recourse models where some of the distributions of the random parameters in the model are known and others are unknown. The key observation in the underlying analysis is that under specific conditions the expected value function of a mixed-integer recourse model allows for a good convex approximation.

5.3.1 Problem formulation

Consider a two-stage mixed-integer recourse model with second-stage value function $v(x, z)$ defined as

$$v(x,z) = \inf_y \left\{ q(\xi)^\top y : Wy = h(\omega) - T(\omega)x, \;\; y \in \mathbb{Z}_+^p \times \mathbb{R}_+^{n_2-p} \right\},$$

where the vector $z = (\xi, \omega)$ represents the random parameters in the model. Similar to (5.1) we assume that q, h, and T are affine functions of these parameters. The distributionally robust mixed-integer recourse model that we consider in this section is

$$\inf_{x \in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + v(x,z)], \tag{5.9}$$

where $\mathcal{P}_z$ represents the (μ, d, β) ambiguity set and $X = \{x \in \mathbb{R}_+^{n_1} : Ax = b\}$. Since $v(x, z)$ is concave in ξ for fixed x and ω, it follows from the same reasoning as in Section 5.2.1 that $\mathbb{P}_{\bar{\xi}}$ defined in Proposition 5.2 is the worst-case distribution of ξ. However, $v(x, z)$ is in general not convex in ω, so that the result of Ben-Tal and Hochman (1972) cannot be applied to derive the worst-case expectations with respect to the distribution of ω. Nevertheless, we are able to use the result if some of the distributions of the random parameters are known and the other distributions are contained in a (μ, d, β) uncertainty set. For ease of exposition we assume in this section that the distribution of the right-hand side random vector $h(\omega)$ is fully known, whereas only limited information is available on the distribution of the technology matrix $T(\omega)$, i.e. $\mathbb{P}_\omega \in \mathcal{P}_\omega$. Moreover, we assume that $h(\omega)$ is independent from $T(\omega)$. Furthermore, since we already discussed the worst-case distribution of ξ we assume that the second-stage costs parameters q are deterministic. For notational convenience, we drop the dependence of h and q on ω and ξ, respectively, and write T as a function of z instead of ω.

Under these assumptions, the distributionally robust mixed-integer recourse model in (5.9) reduces to

$$\inf_{x \in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + Q(x,z)], \tag{5.10}$$

where Q is defined for every realization of z as

$$Q(x,z) = \mathbb{E}_{\mathbb{P}_h}\Big[\inf_y \Big\{q^\top y : Wy = h - T(z)x,\, y \in \mathbb{Z}_+^p \times \mathbb{R}_+^{n_2-p}\Big\}\Big]. \tag{5.11}$$

This expected value function Q is key to solving (5.10), since if Q is convex in z, then we may apply the result of Ben-Tal and Hochman (1972) to obtain the worst-case distribution $\mathbb{P}_{\bar{z}}$ of z. For example, Klein Haneveld et al. (2006) show that under specific conditions on h the expected value function Q of a simple integer recourse model is convex in the tender variables $u := T(z)x$, and thus indeed also convex in z. In general, however, $Q(x,\cdot)$ is not convex, but it may allow for a good convex approximation $\hat{Q}(x,\cdot)$. By replacing Q by $\hat{Q}$ we obtain an approximation of (5.10) for which the objective is convex in z, and thus $\mathbb{P}_{\bar{z}}$ defined in Proposition 5.2 is its worst-case distribution. We derive error bounds on the optimality gaps of the approximating solutions that depend on $\|Q-\hat{Q}\|_\infty$, the maximum difference between Q and $\hat{Q}$.

In Section 5.3.2 we discuss the case where the simple integer recourse function Q is convex in z, and in Section 5.3.3 we derive the error bounds for using convex approximations $\hat{Q}$ for the general two-stage mixed-integer case. In the remainder of this section we briefly survey literature on convex approximations for mixed-integer recourse models and their corresponding error bounds; see also Romeijnders et al. (2014) for an overview.

Klein Haneveld et al. (2006) derived the first error bound for so-called α-approximations of simple integer recourse models which decreases with the total variations $|\Delta|f_i$ of the marginal probability densities functions f_i of the right-hand side random variables h_i. For example for normally distributed random variables this implies that the error bound decreases if the variances of the random variables increase. A similar error bound is obtained by Romeijnders et al. (2015, 2016b) for two different types of convex approximations for totally unimodular integer recourse models. The latter convex approximation is generalized by Romeijnders et al. (2016a) to the general two-stage mixed-integer case. The error bound corresponding to this convex approximation is asymptotic in nature: it converges to zero if all total variations of the probability density functions of the random variables in the model converge to zero.

5.3.2 Simple integer recourse models

The one-sided simple integer recourse model, introduced in Louveaux and van der Vlerk (1993), is a special case of (5.10) for which a closed-from expression for the second-stage value function can be obtained. The expected value function Q is given by

$$Q(x,z) = \sum_{i=1}^m \mathbb{E}_{\mathbb{P}_{h_i}}\Big[q_i \lceil h_i - T_i(z)x \rceil^+\Big], \quad x \in \mathbb{R}^{n_1}, \tag{5.12}$$

where $\lceil s \rceil^+ := \max\{0, \lceil s \rceil\}$, $s \in \mathbb{R}$ and $T_i(z)$ is the i-th row of the matrix $T(z)$. Interestingly, Klein Haneveld et al. (2006) show that this simple integer recourse function Q may be convex in the tender variables $u = T(z)x$, and thus in z, if the underlying random vector h is continuously distributed and every marginal probability density function f_i can be expressed as

$$f_i(s) = H_i(s+1) - H_i(s), \quad s \in \mathbb{R}, \tag{5.13}$$

for some cumulative distribution function H_i with finite mean. This implies that under these conditions the worst-case distribution $\mathbb{P}_{\bar{z}}$ of z can be derived using the results of Ben-Tal and Hochman (1972) (this worst-case distribution is the same for every first-stage decision x).

Proposition 5.5 *Consider the distributionally robust simple integer recourse model*

$$\inf_{x \in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + Q(x,z)], \tag{5.14}$$

where Q is defined in (5.12), $X = \{x \in \mathbb{R}^{n_1}_+ : Ax = b\}$, and the ambiguity set $\mathcal{P}_z$ for the distributions $\mathbb{P}_z$ of z is defined analogously to (5.2). Then, if each random variable h_i has a pdf f_i satisfying (5.13), then the optimization problem in (5.14) is equivalent to

$$\inf_{x \in X} \mathbb{E}_{\mathbb{P}_{\bar{z}}}[c^\top x + Q(x,\bar{z})],$$

where the worst-case distribution $\mathbb{P}_{\bar{z}}$ of z is defined analogously as in Proposition 5.1. □

In case a marginal density function f_i does not satisfy (5.13) a natural approach is to approximate it by a density function $\hat{f}_i$ that is approximately the same as f_i, but does satisfy (5.13), yielding a convex approximation $\hat{Q}$ of Q. This is the main idea behind the so-called α-approximations derived in Klein Haneveld et al. (2006), and their generalization to complete integer recourse models by van der Vlerk (2004). For these convex approximations upper bounds on $\|Q - \hat{Q}\|_\infty$ have been derived. Accordingly, in the next section we assume that a convex approximation $\hat{Q}$ and corresponding upper bound on $\|Q - \hat{Q}\|_\infty$ are available.

5.3.3 Convex approximations

Again consider the general distributionally robust two-stage mixed-integer recourse model defined in (5.10):

$$\eta^* := \inf_{x \in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + Q(x,z)], \tag{5.15}$$

where the expected value function Q, defined in (5.11), is generally non-convex. We assume that Q allows for a good convex approximation $\hat{Q}$ for which $\|Q - \hat{Q}\|_\infty$ is small. Then, we may approximate (5.15) by replacing Q by $\hat{Q}$, obtaining

$$\hat{\eta} := \inf_{x \in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + \hat{Q}(x, z)] \tag{5.16}$$

$$= \inf_{x \in X} \mathbb{E}_{\mathbb{P}_{\bar{z}}}[c^\top x + \hat{Q}(x, \bar{z})], \tag{5.17}$$

where the equality in (5.17) follows from applying the result of Ben-Tal and Hochman (1972) to the convex objective in (5.16). The approximating problem is a convex optimization problem for which the distributions of the random parameters are known. It can be solved efficiently using existing solution methods from SP; see Section 5.4. To guarantee the quality of the approximate solution $\hat{x}$ obtained from solving the optimization problem in (5.17), we derive an error bound on the optimality gap $G(\hat{x}) - \eta^*$, where $G(\hat{x})$ represents the objective value of the solution $\hat{x}$:

$$G(x) := \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}[c^\top x + Q(x, z)], \quad x \in X. \tag{5.18}$$

In fact, we show that $|\hat{\eta} - \eta^*| \leq \|Q - \hat{Q}\|_\infty$ and $G(\hat{x}) - \eta^* \leq 2\|Q - \hat{Q}\|_\infty$; see Theorem 5.1 below.

Interestingly, we may approximate the optimization model in (5.17) by replacing $\hat{Q}$ by the original mixed-integer recourse function Q to obtain the approximating model

$$\tilde{\eta} = \inf_{x \in X} \mathbb{E}_{\mathbb{P}_{\bar{z}}}[c^\top x + Q(x, \bar{z})]. \tag{5.19}$$

This model indirectly approximates the original mixed-integer recourse model (5.15), but it can also be derived directly from (5.15) by assuming that $\mathbb{P}_{\bar{z}}$ is the worst-case distribution in that model. However, using the interpretation of an indirect approximation via the convex approximating model in (5.17), we can derive an error bound for the approximate solution $\tilde{x}$ obtained from solving (5.19).

Theorem 5.1 *Consider the distributionally robust mixed-integer recourse model defined in (5.15) and let $\hat{Q}$ be any convex approximation of the mixed-integer expected value function Q defined in (5.11). Let $\hat{x}$ and $\tilde{x}$ denote optimal solutions of the approximating models defined in (5.17) and (5.19), respectively. Then,*

1. $|\hat{\eta} - \eta^*| \leq \|Q - \hat{Q}\|_\infty$ *and* $G(\hat{x}) - \eta^* \leq 2\|Q - \hat{Q}\|_\infty$,
2. $0 \leq \eta^* - \tilde{\eta} \leq 2\|Q - \hat{Q}\|_\infty$ *and* $G(\tilde{x}) - \eta^* \leq 2\|Q - \hat{Q}\|_\infty$.

Furthermore, since the upper bound on $G(\tilde{x}) - \eta^*$ holds for every convex approximation $\hat{Q}$, it actually holds for the best *convex approximation:*

$$G(\tilde{x}) - \eta^* \leq 2 \inf_{\hat{Q}} \{\|Q - \hat{Q}\|_\infty : \hat{Q} \text{ is convex}\}.$$

Proof. See Appendix 5.A. ■

From a computational point of view, the approximating model in (5.17) is easiest to solve since it is a convex optimization model. The approximating model in (5.19) is a non-convex two-stage mixed-integer recourse model for which the distributions of the random parameters are known. The latter is the main advantage of this approximating model over the original distributionally robust model in which the worst-case distribution of $\mathbb{P}_{\bar{z}}$ still has to be determined and may possibly be different for every first-stage decision x. Nevertheless, solving (5.19) can be a very challenging task. The error bound for this approximating model, however, is the same as for the convex approximating model in (5.17). The fact that the optimality gap of $G(\tilde{x}) - \eta^*$ does not depend on the particular $\hat{Q}$ implies that even if no good convex approximation $\hat{Q}$ of Q is known, we might still approximate the distributionally robust mixed-integer recourse model in (5.15) by assuming that $\mathbb{P}_{\bar{z}}$ is the worst-case distribution of z. If a good convex approximation $\hat{Q}$ of Q is available, then we can use it in the convex approximating model (5.17).

5.4 Stochastic programming models with exponentially many scenarios

In Sections 5.2 and 5.3 we have shown how to reduce a distributionally robust optimization problem to an SP problem for which the distributions of the random variables in the model are known. In particular, in case all decision variables are continuous, we need to solve a continuous stochastic programming model

$$\inf_{x \in X} \Big\{ c^\top x + \mathbb{E}_{\mathbb{P}_{\bar{z}}}[v(x, \bar{z})] \Big\}, \tag{5.20}$$

where the joint distribution of $\bar{z}$ has exponentially many scenarios in the number of random parameters. From a robust optimization point of view this means that the problem in (5.20) is intractable. Indeed, Dyer and Stougie (2006) show that these SP problems are $\#P$-hard. Nevertheless, there has been a vast amount of work in the SP literature that deals with this kind of problems, yielding efficient (approximate) solution methods to these problems, in particular for two-stage problems.

In this section we first discuss so-called simple recourse problems in Section 5.4.1 and we show that for stochastic programming models with such structure, the size of (5.20) does not increase exponentially in the number of random parameters. In Section 5.4.2 we discuss techniques from the SP literature to solve two-stage and multi-stage stochastic programming problems with exponentially many scenarios.

5.4.1 Simple recourse models

In this section we consider so-called simple recourse models introduced by Wets (1983), where the recourse matrix $W = [I_m, -I_m]$ with I_m denoting the m-dimensional

identity matrix. For this model, the second-stage value function is given by

$$v(x,z) = \inf_{y^+,y^-} \Big\{ (q^+(\xi))^\top y^+ + (q^-(\xi))^\top y^- : y^+ - y^- = h(\omega) - T(\omega)x,\, y^+, y^- \in \mathbb{R}^m_+ \Big\},$$

with the conventional indices '+' and '−' representing the surplus and the shortage, respectively. We can obtain an exact expression for this second-stage value function, using among others the *separability of the second-stage problem*:

$$v(x,z) = \sum_{i=1}^{m} \Big(q_i^+(\xi)(h_i(\omega) - T_i(\omega)x)^+ + q_i^-(\xi)(h_i(\omega) - T_i(\omega)x)^- \Big),$$

where $T_i(\omega)$ denotes the i-th row of $T(\omega)$, and $(h_i(\omega) - T_i(\omega)x)^+$ and $(h_i(\omega) - T_i(\omega)x)^-$ denoting the nonnegative and the nonpositive parts of $h_i(\omega) - T_i(\omega)x$, respectively. Suppose that only the right-hand side random vector h is random, then if we drop the dependence of h on z, (5.20) reduces to

$$\inf_{x \in X} \Big\{ c^\top x + \sum_{i=1}^{m} \mathbb{E}_{\mathbb{P}_{\bar{h}_i}} \Big[q_i^+(\bar{h}_i - T_i x)^+ + q_i^-(\bar{h}_i - T_i x)^- \Big] \Big\}, \tag{5.21}$$

where $\mathbb{P}_{\bar{h}_i}$ is the worst-case distribution of h_i as defined in Proposition 5.1. Since it is a three-point distribution, the size of the problem in (5.21) only increases linearly in m. The key observation here is that due to the separability of the second-stage problem the simple recourse model in (5.21) only involves the m marginal distributions $\mathbb{P}_{\bar{h}_i}$, each with three scenarios, instead of the joint distribution $\mathbb{P}_{\bar{h}}$ with 3^m scenarios.

In case there is also uncertainty in the technology matrix T, the simple recourse problem is not completely separable in the random parameters. However, we show in the operating room experiment of Section 5.5.1 that we can use the structure of the problem to substantially speed up the existing algorithms.

5.4.2 Stochastic Programming approaches

The fact that the size of the problem grows exponentially in the number of random parameters is common in SP, and many SP approaches are aimed at reducing the number of scenarios. In this section we survey some relevant SP literature.

One of the most frequently used solution methods is the sample average approximation (SAA), discussed in e.g. Shapiro et al. (2009). The idea of this method is to replace the original worst-case distribution of $\bar{z}$ in (5.20) by a sample z^s, $s = 1, \ldots, N_s$, where N_s is much smaller than the number of scenarios of $\bar{z}$, yielding

$$\inf_{x \in X} \Big\{ c^\top x + \frac{1}{N_s} \sum_{s=1}^{N_s} v(x, z^s) \Big\}. \tag{5.22}$$

If the sample size N_s is small, then the approximation in (5.22) is easier to solve than the original model in (5.20). We may solve (5.22) for several different samples of $\bar{z}$

yielding (possibly) different first-stage solutions x, and use an out-of-sample test to determine the best among them. In the operating room experiment of Section 5.5.1 we show that the SAA method may give near-optimal solutions.

Alternatively, we may use other approaches to reduce the number of scenarios. For example, Dupačová et al. (2003) and Heitsch and Römisch (2003) do so by combining similar scenarios. Pflug (2001) uses the Wasserstein metric to construct a discrete probability distribution (with few scenarios) that minimizes the distance between the original and approximating distribution. His method can also be applied to multi-stage stochastic programming models. Approximations relying on a reduced scenario set are justified by stability results of e.g. Römisch (2003) which shows that a small change in the distributions of the random parameters only result in a small change in the optimal first-stage solutions.

For two-stage stochastic programming models with only a modest number of scenarios efficient solution methods are available. Most of them rely on decomposition of the problem and are variants of the L-shaped algorithm of van Slyke and Wets (1969); see e.g. Ruszczyński (1986) and Higle and Sen (1991) for well-known examples. We refer to Zverovich et al. (2012) for a recent survey comparing several decomposition methods. Although multi-stage stochastic programming models are considerably more difficult to solve than two-stage models, several solution methods do exist. For the interested reader we mention progressive hedging (Rockafellar and Wets 1991), nested Benders' decomposition (Birge 1985), and stochastic dual dynamic programming (Pereira and Pinto 1991).

So far we have only discussed how to obtain a first-stage solution. However, when this solution is obtained by solving an approximation of the original stochastic programming problem, then we may use sampling to assess the quality of the solution; see e.g. the Multiple Replications Procedure (MRP) of Bayraksan and Morton (2009). Different sampling methods, such as Latin Hypercube sampling, may be used to reduce the bias and sample variance of the optimality gap of the approximating solution. We use the MRP to assess the quality of a surgery-to-OR assignment in the operating room experiment of Section 5.5.2.

5.5 Numerical experiments

In this section we present three numerical experiments to illustrate the advantages of the approach developed. The first experiment, a modified version of the operating room (OR) scheduling problem of Denton et al. (2010), illustrates (i) how to reduce the computational effort related to the exponential number of scenarios by using SP techniques and exploiting the problem's properties and (ii) the differences in the performance of distributionally robust solutions compared to other methods used in

OR management.

The second experiment, related also to OR management and involving integer recourse variables shows (i) how the novel theoretical results of Section 5.3 can be used to construct intuitive convex approximations of this integer recourse model, (ii) how to solve it efficiently, and (iii) how to use additional existing techniques to obtain better bounds on the performance of the optimal solutions.

In the third and last experiment, which is a continuation of the inventory management experiment from Postek et al. (2015), we show (i) how our approach is applied to multi-stage problems, (ii) how feasible decisions can be constructed for uncertainty realizations not belonging to the discrete worst-case support, and (iii) we provide managerial insights regarding the value of distributional information and the trade-off between worst-case objective value and worst-case expected objective value.

5.5.1 Operating room scheduling under uncertainty

We apply the method proposed in Section 5.2.1 to the OR scheduling problem introduced by Denton et al. (2010). In this problem, surgeries with random durations have to be assigned to ORs before the durations of these surgeries are known. Fixed costs are incurred for every OR that is opened, and for each OR overtime costs are incurred if the actual total duration of the surgeries exceeds a regular work day of T minutes. Contrary to Denton et al. (2010), we assume that the probability distributions of the surgery durations are (partially) unknown and, hence, we minimize the total worst-case expected costs using the result of this paper. We carry out numerical experiments to show that for problem instances with 10 or 15 surgeries as in Denton et al. (2010), we are able to obtain the optimal surgery-to-OR allocation with reasonable computational effort.

In Section 5.5.1.1 we define the OR scheduling problem and list the various solution methods we use and which are detailed in Appendix 5.B. In Section 5.5.1.2 we carry out the numerical experiments.

5.5.1.1 Problem formulation

The OR scheduling problem can be formulated as a two-stage recourse model, where in the first stage we have to determine how many ORs to open and the assignment of the surgeries to the ORs. With N denoting the number of surgeries that have to be performed, we define y_{ij} for every $i, j = 1, \ldots, N$, as a binary variable equal to 1 if surgery j is assigned to the i-th OR, and 0 otherwise. Thus, we assume that there N ORs available. Accordingly, we define x_i for every $i = 1, \ldots, N$, as a binary variable equal to 1 if the i-th OR is opened, and 0 otherwise. Furthermore, for every opened OR we incur fixed costs c_f and for every minute of overtime exceeding a regular

workday of T minutes we incur variable costs c_v per OR. Let z represent the random vector of surgery durations and θ_i the minutes of overtime in the i-th OR. Then, in case the surgery durations z would be deterministic the OR scheduling problem reads

$$\begin{aligned}
\min_{x,y,\theta} \quad & \sum_{i=1}^{N} c_f x_i + \sum_{i=1}^{N} c_v \theta_i \\
\text{s.t.} \quad & \sum_{i=1}^{N} y_{ij} = 1, && j = 1, \ldots, N, && (5.23)\\
& y_{ij} \leq x_i, && i, j = 1, \ldots, N, && (5.24)\\
& \theta_i \geq \sum_{j=1}^{N} z_j y_{ij} - T x_i, && i = 1, \ldots, N, && (5.25)\\
& x_i \in \{0,1\}, \;\; y_{ij} \in \{0,1\}, \;\; \theta_i \geq 0, && i, j = 1, \ldots, N. && (5.26)
\end{aligned}$$

Constraint (5.23) means that every surgery j is assigned to exactly one OR, constraint (5.24) models that surgery j can only be assigned to the i-th OR if it is opened, and constraint (5.25) defines θ_i as the minutes of overtime for the i-th OR.

We let X denote the set of feasible first-stage decisions x and y satisfying (5.23), (5.24), and (5.26). In addition, we assume that X includes several symmetry breaking constraints introduced in Denton et al. (2010). For example, we assume without loss of generality that $x_1 \geq \cdots \geq x_N$. Moreover, if the surgeries j_1 and j_2 with $j_1 < j_2$ are of the same type, then we assume that surgery j_1 (j_2) is assigned to OR k_1 (k_2), with $k_1 \leq k_2$, respectively:

$$\sum_{k=1}^{i} y_{k,j_1} \geq \sum_{k=1}^{i} y_{k,j_2}, \quad i = 1, \ldots, N.$$

Similar as Denton et al. (2010) we assume that the surgery durations z are random and unknown when the surgery-to-OR assignment has to be made. Contrary to this reference, however, we assume that the distribution $\mathbb{P}_z$ of the random vector z is unknown and belongs to a (μ, d, β) ambiguity set $\mathcal{P}_z$ as defined in Section 5.2.1. The objective is to find a surgery-to-OR assignment, i.e., to determine $(x, y) \in X$, that minimizes the worst-case expected total costs. Given a first-stage decision $(x, y) \in X$ and a realization z of surgery durations, the number of minutes of overtime in the i-th OR is

$$\theta_i(x, y, z) = \Big(\sum_{j=1}^{N} z_j y_{ij} - T x_i \Big)^+. \tag{5.27}$$

The OR scheduling problem minimizing worst-case expected costs is thus given by

$$\min_{(x,y)\in X} \sup_{\mathbb{P}_z \in \mathcal{P}_z} \Big\{ \sum_{i=1}^{N} c_f x_i + c_v \mathbb{E}_{\mathbb{P}_z} \Big[\sum_{i=1}^{N} \Big(\sum_{j=1}^{N} z_j y_{ij} - T x_i \Big)^+ \Big] \Big\}. \tag{5.28}$$

Since the objective function in (5.28) is convex in z for every $(x, y) \in X$, we can use the result of Ben-Tal and Hochman (1972) to obtain the worst-case distribution $\mathbb{P}_{\bar{z}}$ as defined in Proposition 5.1, and thus the optimization problem in (5.28) reduces to

$$\min_{(x,y)\in X} \Big\{ \sum_{i=1}^{N} c_f x_i + c_v \mathbb{E}_{\mathbb{P}_{\bar{z}}} \Big[\sum_{i=1}^{N} \Big(\sum_{j=1}^{N} \bar{z}_j y_{ij} - T x_i \Big)^+ \Big] \Big\}. \tag{5.29}$$

The optimization problem in (5.29) is a two-stage recourse model with binary first-stage variables and continuous second-stage variables, and where the random vector $\bar{z}$ of surgery durations has 3^N scenarios. In fact, this problem has the simple recourse structure discussed in Section 5.4.1. However, contrary to Section 5.4.1, here the randomness is in the technology matrix and not in the right-hand side, so that the number of scenarios does not necessarily reduce to $3N$ as in Section 5.4. Nevertheless, we will use the structure of the problem to deal with exponentially many scenarios.

We use several solution methods to solve the optimization problem in (5.29). The first is a Sample Average Approximation (SAA) method, see Shapiro et al. (2009), which is very easy to implement in practice. The second method (LDR-WCEC: Linear Decision Rules - Worst-Case Expected Cost) uses linear decision rules (LDR) for the overtime costs θ_i so that the optimal surgery-to-OR assignment in this approximating optimization problem can be obtained very fast. The drawback of these two methods is that they only yield an approximate solution to (5.29). Therefore we also use an L-shaped algorithm, see van Slyke and Wets (1969), which yields the optimal solution to (5.29). The challenge of this exact algorithm is to deal with an exponential number of scenarios. In Appendix 5.B, we discuss the SAA and LDR-WCEC methods, and the L-shaped algorithm for this OR experiment. There, we also present several ideas to deal with the exponential number of scenarios, speeding up computations for the L-shaped algorithm considerably. In Section 5.5.1.2 we carry out numerical experiments and find, among others, that the SAA method yields near-optimal solutions within reasonable time limits.

5.5.1.2 Numerical experiments

In this section we carry out numerical experiments on problem instances of similar size as in Denton et al. (2010), i.e. with $N = 10$ and $N = 15$. In all experiments we assume that $c_f = 1$ and $c_v = 0.333$ or $c_v = 0.0833$, similar as in Denton et al. (2010). To obtain the parameters of the (μ, d, β) ambiguity set $\mathcal{P}_z$ we use data on the surgery duration distributions given in Gul et al. (2011). In this reference, estimates of surgery duration distributions are given for several types of surgeries. We use these estimates to compute μ, d, a, b, and β, where a and b represent the 0.1% and 99.9% quantile of the distribution. In Table 5.1 the data of the four types of surgeries that we consider in our experiments are given.

Table 5.1 – OR experiment - parameters of the surgery duration distributions (Gul et al. 2011).

Surgical group	μ	a	b	d	β
Oral Maxillofacial procedure	36.00	2.2	307.2	22.3	0.34
Pain Medicine	20.93	2.3	125.3	10.6	0.37
Ophthalmology	41.63	11.9	125.5	12.6	0.42
Urology	138.16	37.7	433.1	43.3	0.42

For all four combinations of N and c_v we generate 50 problem instances by randomly sampling with equal probabilities N surgery types from Table 5.1. We only report results for $N = 15$, since results for $N = 10$ are similar. For every problem instance we use the SAA method with $N_s = 1000$, the LDR-WCEC method, and the L-shaped algorithm to obtain surgery-to-OR assignments (x, y). In addition, we also obtain $(x, y) \in X$ minimizing the best-case expected costs (min-BCEC) using a similar L-shaped algorithm as for minimizing the worst-case expected costs, and we obtain the surgery-to-OR assignment $(x, y) \in X$ minimizing the worst-case costs (min-WC). For all these first-stage solutions (x, y), we calculate the fixed costs (FC), the worst-case expected costs (WCEC), the best-case expected costs (BCEC), the expected costs (EC), the worst-case costs (WCC) over the support:

$$\mathcal{Z} = [a_1, b_1] \times \ldots \times [a_N, b_N],$$

and the running time (RT) of the algorithm in seconds. Here, the fixed costs (FC) represent the number of opened ORs since $c_f = 1$. Moreover, the expected costs (EC) are estimated using a sample of 100,000 from the surgery duration distributions given in Gul et al. (2011). Furthermore, to facilitate comparison with their results, the worst-case costs (WCC-τ), depending on a parameter τ, are calculated using the same uncertainty set as in Denton et al. (2010):

$$\mathcal{Z} = \left\{ z \in \mathbb{R}_+^N : \quad z_j \in [a_i, b_i] \ \forall j, \quad \sum_{j=1}^{N} \frac{z_j - a_j}{b_j - a_j} \leq \tau \right\}.$$

Here, τ is a parameter representing how many surgeries can attain their maximum duration. The averages of these performance measures over the 50 problem instances are given in Table 5.2.

We conclude from Table 5.2 that the SAA method and the L-shaped algorithm yield very similar results. This implies that, although the SAA solution does not necessarily minimize the worst-case expected costs, its solution is (near-)optimal. Moreover, the surgery-to-OR assignment obtained using linear decision rules for minimizing worst-case expected costs (LDR-WCEC) is more stable since the worst-case costs with $\tau = 2$ and $\tau = 4$ is much smaller. However, in expectation this LDR-WCEC solution is not

Table 5.2 – OR experiment - average values of the performance measures over 50 problem instances with $N = 15$ and $c_v = 0.333$. For every solution, we report the fixed costs (FC), the worst-case expected costs (WCEC), the best-case expected costs (BCEC), the expected costs (EC), the worst-case costs (WCC), and the running time (RT) of the algorithm in seconds.

Solution method	FC	BCEC	EC	WCEC	WCC-1	WCC-2	WCC-4	RT (in seconds)
min-BCEC	2.50	**2.69**	3.11	4.66	3.00	15.49	30.50	8.06
SAA-WCEC	3.28	3.28	3.41	4.27	3.36	14.62	28.89	63.26
L-shaped-WCEC	3.30	3.30	3.42	**4.25**	3.37	14.73	29.30	167.19
LDR-WCEC	8.08	8.08	8.09	8.08	8.08	8.70	10.56	1.40
min-WC	8.98	8.98	8.99	8.98	8.98	8.98	9.12	0.29

good for these problem instances. This is because the number of ORs that are opened in this solution, i.e. the fixed costs (FC), are much larger than for the min-BCEC, SAA, and the L-shaped approaches. We observe in Table 5.2 that the fixed costs are smallest for the min-BCEC solution. This is as expected since the solution minimizes the best-case expectation corresponding to surgery duration distributions for which the longest possible surgery durations are smaller than for the worst-case expectation. Because fewer ORs are opened for this solution, its worst-case expectation is larger than for the L-shaped algorithm and SAA method. On average however, i.e. sampling from the estimated surgery duration distributions of Gul et al. (2011), the min-BCEC solution performs better. Comparing the running times of the algorithms we observe that the LDR-WCEC, min-BCEC and min-WC methods run within several seconds, whereas the SAA method and L-shaped algorithm require on average one minute and almost three minutes, respectively. Given that the SAA method can be implemented more efficiently than in our experiments (using e.g. a decomposition algorithm) and that the L-shaped algorithm minimizes the exact worst-case expected costs under 3^N scenarios, these methods run within reasonable time limits.

Table 5.3 – OR experiment - average values of the performance measures over 50 problem instances with $N = 15$ and $c_v = 0.0833$. Terminology as in Table 5.2.

Solution method	FC	BCEC	EC	WCEC	WCC-1	WCC-2	WCC-4	RT (in seconds)
min-BCEC	2.1	**2.4**	2.53	2.9	2.34	5.51	9.69	4.57
SAA-WCEC	2.08	2.42	2.53	2.89	2.32	5.31	10.2	30.93
L-shaped-WCEC	2.08	2.42	2.53	**2.89**	2.33	5.34	10.2	24.7
LDR-WCEC	3.34	3.4	3.45	3.69	3.35	6.05	9.85	0.65
min-WC	8.44	8.44	8.44	8.44	8.44	8.44	8.65	0.45

In Table 5.3 the results for $N = 15$ and $c_v = 0.0833$ are given. Comparing to the previous case with $c_v = 0.333$, we observe that on average the number of opened ORs is smaller than in Table 5.2 since the per minute overtime costs are smaller. Moreover,

the LDR-WCEC solution does not have as large worst-case expected costs, but the approximation is still not very good in expectation. Furthermore, the difference between the optimal worst-case expected costs and optimal best-case expected costs, displayed in bold face in Tables 5.2 and 5.3, is smaller now. Since this difference yields an upper bound on the *value of distributional information* of the surgery durations, we conclude that for $c_v = 0.333$ we would be willing to spend more time and effort to better estimate the distributions of the surgery durations.

Finally, we report on the efficiency of our tailored implementation of the L-shaped algorithm, as described in Appendix 5.B.3. To illustrate the reductions in the number of scenarios that need to be used by our L-shaped algorithm, we computed the average number of scenarios that had to be evaluated per L-shaped iteration over all 50 runs. The results are given in Table 5.4 and we can see that on average we need between 5% and 15% of the scenarios if the number of surgeries is 10 and less than 1% of the scenarios in case $N = 15$.

Table 5.4 – OR experiment - average numbers of scenarios to evaluate per iteration in the L-shaped algorithm.

Experiment	N	c_v	Evaluations	Evaluations/$3^N \times 100\%$
1	10	0.333	2833	4.80
2	10	0.0833	7146	12.10
3	15	0.333	8509	0.059
4	15	0.0833	25685	0.179

5.5.2 Numerical experiments for two-stage ambiguous integer recourse models

In this section we again consider the OR scheduling problem of Section 5.5.1. However, here we assume that overtime wages are paid in full hours: if the overtime in a given OR is only a few minutes, then still its OR staff has to be paid a full hour of overtime work. In addition we assume that there is uncertainty in the regular work day duration T. This duration may be interpreted as the effective time spent on performing surgeries and may be smaller (or larger) than the targeted 480 minutes due to inefficiency (or efficiency) of the OR staff.

In Section 5.5.2.1 we show that this problem can be modelled as a distributional robust integer recourse model of Section 5.3, and we derive a convex approximation for this problem. In Section 5.5.2.2 we evaluate this convex approximation using numerical experiments.

5.5.2.1 Problem definition and convex approximation

Since we do not have detailed information about the efficiency of the OR staff, we assume that for every OR i the duration of a regular work day equals $T+\epsilon_i$, where the probability distribution $\mathbb{P}_{\epsilon_i}$ of ϵ_i belongs to a (μ, d, β)-uncertainty set with $E[\epsilon_i] = 0$. Moreover, contrary to Section 5.5.1, we assume that the probability distributions of the surgery durations z are known, for example based on historical data. Under these assumptions, the problem can be cast into the framework of Section 5.3, where the distributions of the surgery durations z are known and the distributions of the work day durations $(T + \epsilon_i)x_i$ are contained in a $(\mu, d, \underline{\ })$ ambiguity set. Again, letting X denote all feasible surgery-to-OR assignments (x, y), the optimization problem we consider is given by

$$\inf_{(x,y)\in X}\left\{\sum_{i=1}^{N} c_f x_i + \sup_{\mathbb{P}_\epsilon\in\mathcal{P}_{(\mu,d)}}\left\{\mathbb{E}_{\mathbb{P}_\epsilon}\left[\mathbb{E}_{\mathbb{P}_z}\left[\sum_{i=1}^{N} 60c_v\left\lceil\Big(\sum_{j=1}^{N} z_j y_{ij} - (T+\epsilon_i)x_i\Big)/60\right\rceil^+\right]\right]\right\}\right\}.$$

Here, all durations are in minutes and the round-up operator ensures that overtime wages are paid in full hours.

Because of the round-up operator, the objective function is not convex in ϵ and thus the results of Ben-Tal and Hochman (1972) cannot be applied. This means that we do not know the worst case distribution of $\mathbb{P}_\epsilon$. In fact, the worst-case distribution may be different for every surgery-to-OR assignment $(x, y) \in X$. Following Section 5.3, we define the expected value function Q as

$$Q(x, y, \epsilon) = \mathbb{E}_{\mathbb{P}_z}\left[\sum_{i=1}^{N} 60c_v\left\lceil\Big(\sum_{j=1}^{N} z_j y_{ij} - (T+\epsilon_i)x_i\Big)/60\right\rceil^+\right],$$

and we consider its convex approximation

$$\hat{Q}(x, y, \epsilon) = \mathbb{E}_{\mathbb{P}_z}\left[\sum_{i=1}^{N} c_v\Big(\sum_{j=1}^{N} z_j y_{ij} - (T+\epsilon_i-30)x_i\Big)^+\right].$$

Here, we simultaneously relax the integrality of the overtime hours and subtract 30 minutes from the work day duration (if the i-th OR is opened). The rationale of doing so is that on average we have to pay approximately 30 minutes of additional overtime if overtime is paid in full hours.

For the convex approximating model with Q replaced by $\hat{Q}$ we can apply the results from Section 5.3 to conclude that the worst-case distribution equals $\mathbb{P}_{\bar{\epsilon}}$ for every $(x, y) \in X$. The approximating model becomes

$$\begin{aligned}&\inf_{(x,y)\in X}\left\{\sum_{i=1}^{N} c_f x_i + \mathbb{E}_{\mathbb{P}_{\bar{\epsilon}}}[\hat{Q}(x, y, \bar{\epsilon})]\right\}\\ =\ &\inf_{(x,y)\in X}\left\{\sum_{i=1}^{N} c_f x_i + \mathbb{E}_{\mathbb{P}_{\bar{\epsilon}}}\left[\mathbb{E}_{\mathbb{P}_z}\left[\sum_{i=1}^{N} c_v\Big(\sum_{j=1}^{N} z_j y_{ij} - (T+\epsilon_i-30)x_i\Big)^+\right]\right]\right. \qquad (5.30)\end{aligned}$$

This model can be solved e.g. using SAA yielding an approximating surgery-to-OR assignment $(\hat{x}, \hat{y})$.

5.5.2.2 Numerical experiments

For our numerical experiments we consider the same setting as in Section 5.5.1. That is, the fixed costs for opening an OR are normalized, i.e. $c_f = 1$, and we consider two different cases for the overtime costs: $c_v = 0.0333$ and $c_v = 0.00833$. Moreover, we only consider the types of surgeries presented in Table 5.1. For every OR i, we assume that the ambiguity set of $\mathbb{P}_{\epsilon_i}$ is defined by $a_i = -60$, $b_i = 60$, $\mu_i = 0$, $d_i = 30$, and $\beta_i = 0.5$. This means that the regular work day duration $T + \epsilon_i$ will be between 420 and 540 minutes (i.e. 7 and 9 hours).

To solve the convex approximating model in (5.30) we use SAA with sample size $\hat{N}_s$ to obtain an approximating surgery-to-OR assignment. We repeat this procedure ten times, obtaining ten possibly different surgery-to-OR assignments, and use an out-of-sample test of size 10,000 to obtain the best among them. We let $(\hat{x}, \hat{y})$ denote this surgery-to-OR assignment. Contrary to Section 5.5.1 we are not able to determine the optimal surgery-to-OR assignment. That is why we have to use a different approach to determine the quality of the solution $(\hat{x}, \hat{y})$. It turns out that using the maximum difference between the expected value function Q and its convex approximation $\hat{Q}$, as suggested in Theorem 5.1 yields error bounds that are too large for this particular problem. That is why we instead use a combination of the Multiple Replications Procedure (MRP) discussed in e.g. Bayraksan and Morton (2009) and total variation error bounds. In Appendix 5.C.1 we discuss this approach in more detail. The result is an (approximate) 95% confidence interval on the optimality gap of $(\hat{x}, \hat{y})$.

Table 5.5 – Integer OR experiment - numerical results for the integer OR problem with $N = 15$ over 10 problem instances, where the surgery types are randomly generated based on Table 5.1, and $c_v = 0.0333$. Here, $\hat{N}_s$ denotes the sample size used to obtain the approximating solution, and FC denotes the fixed costs of this solution. Next, ELB OBJ VAL gives an expected lower bound on the optimal objected value, and next we have an (approximate) 95% confidence interval on the absolute optimality gap and an upper bound on the relative optimality gap. Finally, RT denotes the average running time in seconds of solving the SAA of the convex approximating model with a sample size of $\hat{N}_s$.

$\hat{N}_s$	FC	ELB OBJ VAL	95% CI OPT GAP	REL OPT GAP	RT (in sec)
10	2.8	3.40	[0,0.214]	6.38%	1.16
100	3.0	3.40	[0,0.086]	2.52%	11.5
1000	3.0	3.40	[0,0.067]	1.98%	143.5

The results in Tables 5.5 and 5.6 are obtained by solving 10 problem instances with $c_v = 0.0333$ and $c_v = 0.00833$, respectively, and $N = 15$ surgeries, each randomly selected from Table 5.1. We obtained three approximating surgery-to-OR assignment by solving the SAA of (5.30) with sample size $\hat{N}_s = 10, 100$, and 1000. As can be

Table 5.6 – Integer OR experiment - numerical results for the integer OR problem with $N = 15$ over 10 problem instances, where the surgery types are randomly generated based on Table 5.1, and $c_v = 0.00833$. Terminology the same as in Table 5.5.

$\hat{N}_s$	FC	ELB OBJ VAL	95% CI OPT GAP	REL OPT GAP	RT (in sec)
10	2.8	3.41	[0,0.243]	6.68%	1.43
100	2.9	3.41	[0,0.093]	2.72%	9.98
1000	3.0	3.41	[0,0.074]	2.20%	176.5

observed in the tables, the computation time required to obtain these surgery-to-OR assignments increases in the sample size $\hat{N}_s$. As expected, the quality of the surgery-to-OR assignments also increases in the sample size since the relative optimality gap (REL OPT GAP) decreases from approximately 6% for $\hat{N}_s = 10$ to 2% for $\hat{N}_s = 1000$. This difference is not caused by the number of ORs that is opened in the two different cases but rather by the different surgery-to-OR assignments. Overall we conclude that the 95% confidence intervals on the optimality gap are surprisingly small compared to the expected lower bound on the objective value for $\hat{N}_s = 1000$, in particular since these distributionally robust integer problems are extremely hard to solve and we are not able to calculate the optimal solution. We would like to stress that the values in Tables 5.5 and 5.6 are upper bounds on the optimality gap that hold with high probability. The actual value of the optimality gaps might be even smaller. In Appendix 5.C.1 more details can be found on how the confidence interval on the optimality gap is obtained.

5.5.3 Inventory experiment

5.5.3.1 Introduction

Our final experiment concerns a multi-stage problem - an inventory management example adapted from Ben-Tal et al. (2005), used also in Postek et al. (2015), comprising a single product with inventory managed over T stages. At the beginning of each stage t the decision maker has an inventory of size I_t and he orders a quantity x_t for unit price c_t. The customers then place their demands z_t. The retailer's status at the beginning of the planning horizon is given by the parameter I_1 (initial inventory). Apart from the ordering cost, the following costs are incurred over the planning horizon: (i) holding cost $h_t(I_{t+1})^+$, where h_t is the unit holding cost, (ii) shortage cost $s_t(-I_{t+1})^+$, where s_t is the unit shortage cost.

Inventory I_{t+1} left at the end of stage T has a unit salvage value s. Also, one must impose $h_t - s \geq -s_t$ to maintain the problem's convexity. A practical interpretation of this constraint is that in the last stage it is more profitable to satisfy the customer demand rather than to be left with an excessive amount of inventory. The constraints

Table 5.7 – Inventory experiment - ranges for parameter sampling in the inventory experiment.

Parameter	Range	Parameter	Range	Parameter	Range
a_t	$[0, 20]$	L_t	0	$\hat{U}_t$	$0.8\sum_{t=1}^{T} U_t$
b_t	$[a_t, a_t + 100]$	U_t	$[50, 70]$	h_t	$[0, 5]$
c_t, s_t	$[0, 10]$	$\hat{L}_t$	0	I_1	$[20, 50]$

in the model include (i) balance equations linking the inventory in each stage to the inventory, order quantity, and demand in the preceding stage, (ii) upper and lower bounds on the order quantities in each stage $L_t \leq x_t \leq U_t$, (iii) upper and lower bounds on the cumulative order quantity up to stage $\widehat{L}_t \leq \sum_{\tau=1}^{t} x_\tau \leq \widehat{U}_t$.

The problem to solve without uncertainty in the demand is

$$\begin{aligned} \min_{x} \quad & \sum_{t=1}^{T} \{c_t x_t + h_t (I_{t+1})^+ + s_t (-I_{t+1})^+\} & \\ \text{s.t.} \quad & I_{t+1} = I_t + x_t - z_t, & t = 1, \ldots, T \\ & L_t \leq x_t \leq U_t, & t = 1, \ldots, T \\ & \widehat{L}_t \leq \sum_{\tau=1}^{t} x_\tau \leq \widehat{U}_t, & t = 1, \ldots, T. \end{aligned} \tag{5.31}$$

To model uncertainty about demands $z = (z_1, \ldots, z_T)$, we assume that $\mathcal{Z}$ is the support defined as $\mathcal{Z} = \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_t$, where $\mathcal{Z}_t = [a_t, b_t]$, $t = 1, \ldots, T$, which corresponds to z being a random variable with independent components. For the ambiguity set of the uncertain demand distribution, we set $\mu_t = (a_t + b_t)/2$, $\mathbb{E}_{\mathbb{P}}|z_t - \mu_t| = (b_t - a_t)/4$, and $\mathbb{P}(z_t \geq \mu_t) = \beta$ for $t = 1, \ldots, T$. We use the same 50 problem instances with $T = 6$ as Postek et al. (2015); the ranges for the uniform sampling of parameters are given in Table 5.7.

Our goal is to obtain and compare decisions corresponding to various solution approaches for this multi-stage problem with distributional uncertainty. Among others, there are two questions related to a multi-stage problem in such a setting: (i) what should be the minimized objective criterion?; and (ii) how to make the later-stage decisions adjust to the observed demands?

With respect to the first question, we propose to minimize either the worst-case expectation as the most pessimistic value in our setting, or the best-case expectation as the most optimistic one. The closed-form formulation of the worst-case and best-case expectations of the objective function in (5.31) can be evaluated using the methodology of Section 5.2.2.

The second issue – adjustability of decisions – is important as (i) in this way the later-stage decisions can be better for each outcome of the uncertainty, (ii) the best here-and-now decisions can be different if later-stage adjustability is accounted for.

Adjustability is typically achieved by formulating the later-stage decisions as functions (*decision rules*) of the realized demand and then, optimizing the parameters of these functions as decision variables

A simple and common choice for the decision rules is to define them as linear functions $x_{t+1} := x_{n+1,0} + \sum_{j=1}^{t} x_{n+1,j} z_j$, where $t = 1, \ldots, T$, of the observed uncertainties. However, as affinity may be too restrictive, we propose also piecewise-linear decision rules obtained by interpolating the decisions in the finite worst-case (or bast-case) support, described in detail in Appendix 5.D.1. If the coefficients of the decision rules are determined before the planning horizon and they are not altered later, we denote this approach by *evaluation*.

Alternatively to fixing the decision rules and not changing them later, it is possible, having found the optimal solution and implemented the initial decision x_1 and observed z_1, to solve a new optimization problem where new decisions (and decision rules) for stages $2, \ldots, T$ are determined. In contrast to the *evaluation approach*, we denote it as *reoptimization*.

In this setting, we consider three solutions:

- minimizing the worst-case expectation with linear decision rules (results are taken from Postek et al. (2015)). We denote this approach as L-WCE ('L' stands for 'linear' and 'WCE' for 'worst-case expectation');

- minimizing the worst-case expectation using the piecewise-linear decision rules of Appendix 5.D.1. We denote this approach as PL-WCE ('PL' stands for 'piecewise-linear');

- minimizing the best-case expectation using the piecewise-linear decision rules of Appendix 5.D.1. We denote this approach as PL-BCE-β ('BCE' for 'best-case expectation' and β is the skewness parameter for each t as defined in (5.2), which we assume to be the same for all $t = 1, \ldots, T$).

5.5.3.2 Intervals for the expected value of total cost

Due to the distributional uncertainty in our problem, it is not possible to know the exact value of the expected total cost incurred over the planning horizon. However, due to the convexity of the objective function in z it is possible to evaluate both the worst-case and the best-case expectations of the total cost, which gives an interval for the expectation of the objective. Such intervals allow us to compare the three solutions with respect to (i) optimality: minimizing expected costs, lower values are preferable, (ii) range: the narrower an interval, the less ambiguity about the 'true' expected cost.

Table 5.8 – Inventory experiment - evaluation intervals - ranges for the expectation of the objective over $\mathcal{P}_z$ ('expectation range' is computed for a given solution using the upper and lower bound results of Ben-Tal and Hochman (1972) under given assumptions) and worst-case cost ('worst-case value' is the maximum total cost obtained for the single worst-case scenario out of $\mathcal{Z}$). All numbers are averages over the 50 instances.

Objective type	β	Solution				
		L-WCE	PL-WCE	PL-BCE-0.25	PL-BCE-0.5	PL-BCE-0.75
Expectation range	0.25	[1004, 1049]	[973, **1007**]	[**940**, 1178]	[976, 1133]	[1087, 1228]
Expectation range	0.5	[970, 1049]	[943, **1007**]	[1009, 1178]	[**908**, 1133]	[978, 1228]
Expectation range	0.75	[994, 1049]	[960, **1007**]	[1157, 1178]	[986, 1133]	[**905**, 1228]
Worst-case value	-	2384	2358	2628	2553	3005

Intervals for solutions in the evaluation approach. Table 5.8 presents the results on the performance of the three solutions. The PL-WCE solution achieves a better worst-case (maximum total cost over the entire support) objective value (2358 versus 2384), and for each β the upper and lower endpoints of the interval for PL-WCE are smaller than upper and lower endpoints of the interval for L-WCE, for example, $[943, 1007]$ versus $[970, 1049]$ for $\beta = 0.5$. This provides strong evidence that restricting the decision rules to linear functions can have a negative effect on the quality of the solution as measured by the objective function.

We now compare the widths of the intervals corresponding to different solutions, which is our proxy for the value of distributional information and the 'riskiness' of each solution. We observe that PL-BCE solutions give expectation intervals that are overall much more dispersed than the PL-WCE solutions, compare e.g. $[908, 1133]$ (width 225) and $[943, 1007]$ (width 64) in the third row. On average, the intervals corresponding to the PL-BCE solutions are 5 times wider than the ones from PL-WCE solutions. This indicates that minimization of the worst-case expectation (pessimistic approach) may have a 'compressing' impact on the expectation interval, whereas the solutions obtained by minimizing the best-case expectation (optimistic approach) come with a much wider range.

With respect to the value of distributional information (VDI), we can approximate it as follows on the example of the PL-WCE solution. The width of the interval is for $\beta = 0.5$ is given by $1007 - 943 = 64$ which is the VDI. This value, divided by the upper bound on the worst-case performance yields $64/1007 \times 100\% \approx 6.35\%$ – it is the remaining relative uncertainty about the objective expectation. It is questionable whether profits can be gained by knowing or gathering exact data on the distribution since (i) computational handling of this extra information in the optimization problem would be significantly more complicated, (ii) the resulting more precise expectation value would be much more sensitive to estimation errors.

Intervals for solution values in the reoptimization approach. We also con-

Table 5.9 – Inventory experiment - reoptimization intervals. All numbers are averages over the 50 instances. Terminology as in Table 5.8.

Objective type	β	Solution				
		L-WCE	PL-WCE	PL-BCE-0.25	PL-BCE-0.5	PL-BCE-0.75
Expectation range	0.25	[972, 1011]	[965, **1007**]	[**940**, 1040]	[975, 1037]	[995, 1038]
Expectation range	0.5	[941, 1011]	[933, **1007**]	[938, 1040]	[**903**, 1037]	[933, 1038]
Expectation range	0.75	[955, 1011]	[952, **1007**]	[982, 1040]	[942, 1037]	[**905**, 1038]

sider the intervals for the objective function value assuming that the decision maker can reoptimize the solution over time, described in more detail in Appendix 5.D.2. The results are given in Table 5.9. Compared to Table 5.8, it is clear that for each solution and each value of β the corresponding upper and lower bounds of the intervals are not larger than the ones from Table 5.8 (compare for example the lower endpoints for L-WCE). Partly due to this change, the intervals obtained for various solutions become more similar.

5.5.3.3 Simulation study

Apart from simply knowing the intervals to which the expected total costs are guaranteed to belong, one may be interested in the performance of the three solutions in a 'reasonable' simulation setting. Since we do not know the exact distributions of the uncertain random parameters, we use the following two distributions to sample from:

- uniform sample: demand scenarios $\hat{z}$ are sampled from a uniform distribution on the support $\mathcal{Z}$;
- (μ, d) sample: demand scenarios $\hat{z}$ are sampled from a randomly sampled distribution $\hat{\mathbb{P}} \in \mathcal{P}_z$ – the details of the sampling methodology are given in Appendix 5.D.3.

As the (μ, d) sample involves the distributional uncertainty, it is 'broader' than the uniform sample, i.e. it encompasses more than one possible choice for the probability distribution out of the given ambiguity set.

Evaluation. Table 5.10 presents the results in the evaluation approach. The PL-WCE solution again gives better values than L-WCE, both in terms of the mean values - an improvement of 2.81% on the uniform sample, and the standard deviations of the objective function value - for example, an improvement of 3.01% on the (μ, d) sample. Interesting results are also obtained for the PL-BCE solutions: on the (μ, d) sample they perform better than the L-WCE and PL-WCE solutions, despite their focus on the best-case expectation. Also, the PL-BCE solutions provide

Table 5.10 – Inventory experiment - evaluation simulation results. Numbers in brackets denote the % change compared to the L-WCE solution. All numbers are averages over all 50 problem instances.

Value	Sample	Solution				
		L-WCE	PL-WCE	PL-BCE-0.25	PL-BCE-0.5	PL-BCE-0.75
Mean	Uniform	994	**966** (**−2.81**%)	1019 (+2.54%)	999 (+0.54%)	1026 (+3.29%)
Standard deviation	Uniform	259	**251** (**−3.08**%)	272 (+5.09%)	255 (−1.45%)	286 (10.49%)
Mean	(μ, d)	1003	971 (−3.19%)	976 (−2.66%)	**962** (**−4.00**%)	986 (−1.60%)
Standard deviation	(μ, d)	265	257 (−3.01%)	241 (−8.83%)	**223** (**−15.62**%)	248 (−6.40%)

Table 5.11 – Inventory experiment - reoptimization simulation results. Numbers in brackets denote the % change compared to the L-WCE solution. All numbers are averages over all 50 instances.

Value	Solution				
	L-WCE	PL-WCE	PL-BCE-0.25	PL-BCE-0.5	PL-BCE-0.75
Mean value	977.8	**972.8** (**−0.5**%)	986.6 (+0.90%)	974.5 (−0.34%)	980.6 (+0.29%)
Standard deviation	251.2	257.5 (+2.51%)	272.6 (+8.51%)	237.7 (−5.34%)	**236.0** (**−6.03**%)

substantial decreases in the standard deviation of the estimator of the expected total costs compared to the other solutions.

Reoptimization. We also consider the reoptimization version of our experiment where we use 500 demand samples from the (μ, d) distributions (we do not report on the results for the uniform sample as they are nearly the same). Table 5.11 presents the results. The means of the simulated total cost are almost the same for all solutions, differing by less than 1%. This small difference is in line with the results of the previous subsection where the intervals in the reoptimization experiment turn out to be similar as well.

Choosing the right solution. The results of this and the previous subsection give rise to the question whether any of the three solutions is preferable to others. We suggest that this choice depends on three factors: (i) risk-aversion of the decision maker, (ii) amount of computational power available, and (iii) possibility (or not) to re-optimize.

For a risk-averse decision maker, the PL-WCE solution is more flexible (the implied decision rule is piecewise-linear instead of linear) than the L-WCE and gives better worst-case expected performance in evaluation settings, as shown in Table 5.8. On the other hand, it requires a larger computational effort (second criterion) as each worst-case demand scenario requires a separate ordering trajectory. The number of optimization variables in the problem with L-WCE decisions equals $T(T+1)/2$ (one variable for time 1 decisions, 2 for time 2 decisions, etc.) whereas for PL-WCE this number equals $(3^T-1)/2$ (enumerating all the 3^T trajectories and elimination of some double-counted decisions through the *nonanticipativity constraints*). With respect to the third criterion, if reoptimization is possible, then we see that the differences

between the three solutions are very small.

5.5.3.4 Pareto stripe

A common tool in decision support is the Pareto curve, illustrating a tradeoff between two criteria. It is obtained by finding, for fixed bounds on one objective (for example, the worst-case cost over the entire demand support), the minimum of another objective (for example, the mean cost). A strong feature of our approach is that it allows to evaluate both the worst-case and the best-case expectation of a convex function. That is, for a given bound on the worst-case value of the total cost, we are able to identify the entire interval for the expected cost. This gives rise to an extension of the Pareto curve, denoted as the *Pareto stripe*, which in our case depicts how a bound on the worst-case cost affects (i) the best(worst)-case expectation, (ii) the value of distributional information, as measured by the width of this interval.

Mathematically, the Pareto stripe is obtained by minimizing, for a given (fixed) upper bound $C \in \mathbb{R}$ on the worst-case value of the objective function:

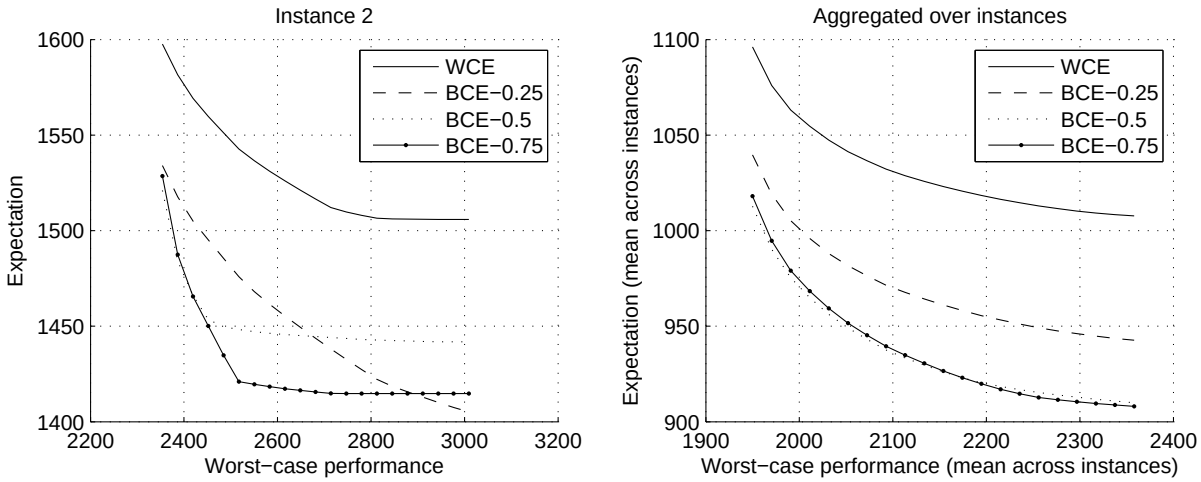
$$\sup_{z \in \mathcal{Z}} \left\{ \sum_{t=1}^{T} \left(c_t x_t + h_t (I_{t+1})^+ + s_t (-I_{t+1})^+ \right) \right\} \leq C, \tag{5.32}$$

the worst-case expectation (or the best-case expectation) by means of the PL-WCE solution (PL-BCE solution, respectively, with different possible values for the skewness β).

The left panel in Figure 5.2 presents such a stripe for a single problem instance. An interesting feature is that the best-case expectations obtained for various values of parameter β need not preserve any monotonicity relation. For example, the best-case expectation when $\beta = 0.75$, obtained for the worst-case bound C of around 2400 (horizontal axis) is the highest of all, but is smaller than best-case expectations for $\beta = 0.25$ and $\beta = 0.5$ when the worst-case bound C is 3000.

The right panel in Figure 5.2 presents the Pareto stripe aggregated over all 50 problem instances. For orientation, we note that the rightmost value of the continuous black curve in the right panel of Figure 5.2 corresponds to the first row of Table 5.8, whereas the rightmost value of the horizontal axis is the worst-case objective value of PL-WCE solutions from Table 5.8. Figure 5.2 provides an assessment of the value of distributional information. We observe that as the bound on the worst-case performance (horizontal axis) grows, the width of the Pareto stripe increases slightly, corresponding to a growth in the VDI (width of the interval compared to its upper bound value) from about 7% to about 11%.

Figure 5.2 – Inventory experiment - upper (WCE) and lower (BCE) bounds on the expectation of the objective function values for a single problem instance (left panel) and aggregated over all 50 instances (right panel). The best-case expectations (lower bounds) have been computed for $\beta \in \{0.25, 0.5, 0.75\}$.



5.6 Conclusion

In this paper we have considered stochastic programming problems with distributional ambiguity. We have shown that under mean - MAD distributional information, the problem admits a closed form reformulation as the corresponding worst-case distributions consist of 3 points per component and is independent from the first stage decisions. This holds both for two-stage and multi-stage continuous models. We have proposed methods to deal with the exponential number of scenarios that perform well in the numerical experiments. For two-stage problems with integer recourse variables, we show how good convex approximations can be derived that have a provable performance guarantee. Our numerical experiments entailing operating room scheduling and inventory management provide also simple yet powerful managerial insights such as (i) the easy-to-calculate value of distributional information (difference between the worst- and best-case expectation under the given information) and (ii) the Parieto stripe, which shows how the interval containing the true expected objective function changes relative to a bound on a certain performance measure. Overall, we have proposed a practical framework of solving a wide class of problems that can easily be implemented in a variety of real-world applications.

Appendices

5.A Proofs

Proof.[Proposition 5.1] For simplicity, assume first that $n_\omega = 1$ and $n_\xi = 0$. Since $v(x, z)$ is a convex function of z then by result of Ben-Tal and Hochman (1972) we

have that:

$$\sup_{\mathbb{P}_{z_1}\in\mathcal{P}_{z_1}} \mathbb{E}_{\mathbb{P}_{z_1}} v(x,z_1) = \tag{5.33}$$
$$= \frac{d_1}{2(\mu_1-a_1)}v(x,a_1) + \left(1 - \frac{d_1}{2(\mu_1-a_1)} - \frac{d_1}{2(b_1-\mu_1)}\right)v(x,\mu_1) + \frac{d_1}{2(b_1-\mu_1)}v(x,b_1),$$

that is, the worst-case expectation of $v(x,z)$ is achieved by a three-point distribution with support $\{a_1,\mu_1,b_1\}$ and probabilities $d_i/2(\mu_i-a_i)$, $1-d_i/2(\mu_i-a_i)-d_i/2(b_i-\mu_i)$ and $d_i/2(b_i-\mu_i)$, respectively. For $n_z \geq 2$ we observe that due to the independence of z_i's we have:

$$\mathcal{P}_z = \mathcal{P}_{z_1} \times \ldots \times \mathcal{P}_{z_{n_\omega}},$$

where

$$\mathcal{P}_{z_i} = \{\mathbb{P}_{z_i} : \quad \text{supp}(z_i) \subseteq [a_i,b_i], \quad \mathbb{E}_{\mathbb{P}}(z_i) = \mu, \quad \mathbb{E}_{\mathbb{P}}\left|z_i-\mu_i\right| = d_i\}, \quad i = 1,\ldots,n_\omega.$$

Therefore, we can apply formula (5.34) component-wise w.r.t. z:

$$\sup_{\mathbb{P}_z\in\mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z} v(x,z) = \sup_{\mathbb{P}_{z_1}\in\mathcal{P}_{z_1}} \mathbb{E}_{\mathbb{P}_{z_1}} \left\{ \sup_{\mathbb{P}_{z_2}\in\mathcal{P}_{z_2}} \mathbb{E}_{\mathbb{P}_{z_2}} \left\{ \ldots \sup_{\mathbb{P}_{z_{n_z}}\in\mathcal{P}_{z_{n_\omega}}} \mathbb{E}_{\mathbb{P}_{z_{n_\omega}}} v(x,(z_1,\ldots,z_{n_\omega}))\ldots \right\}\right\}.$$

Therefore, the support of the worst-case distribution of z is a product of the worst-case distributions of z_i, equal to $\overline{\mathcal{Z}}$, and the probability of a single z^k is equal to the product of the worst-case probabilities of the respective components of z^k, as defined in Proposition 5.1. A similar argument holds for the worst-case expectation w.r.t. ξ and since it is assumed that the components of ω and ξ are mutually independent, the claim follows. ■

Proof.[Proposition 5.2.] The proof is analogous to Proposition 5.1, therefore, we only consider the case $n_\omega = 1$ and $n_\xi = 0$. Since $v(x,z)$ is a convex function of z then by result of Ben-Tal and Hochman (1972) we have that:

$$\inf_{\mathbb{P}_{z_1}\in\mathcal{P}_{z_1}} \mathbb{E}_{\mathbb{P}_{z_1}} v(x,z_1) = \tag{5.34}$$
$$= (1-\beta_1)v(x,\mu_1 - d_1/2(1-\beta_1)) + \beta_1 v(x,\mu_1 + d_1/2\beta_1),$$

that is, the best-case expectation of $v(x,z)$ is achieved by a two-point distribution with support $\{\mu_1 - d_1/2(1-\beta_1), \mu_i + d_1/2\beta_1\}$ and probabilities $(1-\beta_i)$ and β_i, respectively. ■

Proof.[Proposition 5.3.] Consider first problem (5.6) solved at time T. The problem is linear, hence convex in x_T and by Fiacco and Kyparisis (1986) it holds that the optimal value of (5.6) is convex in x_{T-1} and z_{T-1}. Next, consider the problem to

solve at time $T-1$:

$$\inf_{x_{T-1}} \Big\{ c_{T-1}^T x_{T-1} + \sup_{\mathbb{P}_{z_{T-1}} \in \mathcal{P}_{z_{T-1}}} \mathbb{E}_{\mathbb{P}_{z_{T-1}}} v_T(x_{T-1}) : \sum_{s=1}^{T-2} A_{T-1s}(z_{T-2}) x_s + A_{T-1T-1} x_{T-1} = b_{T-1}(z_{T-2}) \Big\}. \tag{5.35}$$

Since $v_T(x_{T-1}, z_{T-1})$ is convex in x_{T-1}, the objective function in (5.35) is also convex in x_{T-1} and, since the remaining constraints are linear in x_{T-1}, the problem is convex in x_{T-1}. Again, by Fiacco and Kyparisis (1986) it holds that $v_{T-1}(x_{T-2}, z_{T-2})$ is convex in z_{T-2} and x_{T-2}. The same argument is applied recursively to time stages $T-1, T-2, \ldots, 2$ which proves the claim. ■

Proof.[Theorem 5.1.] Let x^* denote an optimal solution to the optimization problem in (5.10). Then,

$$\begin{aligned} \eta^* \leq G(\hat{x}) &= \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z} \Big[c^T \hat{x} + Q(\hat{x}, z) \Big] \Big\} \\ &\leq \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z} \Big[c^T \hat{x} + \hat{Q}(\hat{x}, z) \Big] \Big\} + \|Q - \hat{Q}\|\infty \\ &= \hat{\eta} + \|Q - \hat{Q}\|_\infty. \end{aligned}$$

Here, the first inequality holds since $\hat{x}$ is not necessarily optimal in the original model. Similarly, we have

$$\begin{aligned} \hat{\eta} &\leq \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z} \Big[c^T x^* + \hat{Q}(x^*, z) \Big] \Big\} \\ &\leq \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z} \Big[c^T x^* + \hat{Q}(x^*, z) \Big] \Big\} + \|Q - \hat{Q}\|\infty \\ &= \eta^* + \|Q - \hat{Q}\|_\infty. \end{aligned}$$

Combining $\eta^* \leq \hat{\eta} + \|Q - \hat{Q}\|_\infty$ and $\hat{\eta} \leq \eta^* + \|Q - \hat{Q}\|_\infty$ yields the first inequality in (i). Furthermore, using $G(\hat{x}) \leq \hat{\eta} + \|Q - \hat{Q}\|_\infty$ and $\hat{\eta} \leq \eta^* + \|Q - \hat{Q}\|_\infty$, it follows that

$$G(\hat{x}) \leq \eta^* + 2\|Q - \hat{Q}\|_\infty,$$

and from this the second inequality in (i) follows immediately.

Further, observe that $\tilde{\eta}$ is a lower bound for η^* since $\mathbb{P}_{\tilde{z}}$ is not necessarily the worst-case distribution in model (5.10), and thus $0 \leq \eta^* - \tilde{\eta}$. Next, let $\hat{Q}$ be a convex

approximation of Q. Then, the remaining inequalities in (ii) follow directly from

$$\begin{aligned}
\eta^* \leq G(\tilde{x}) &= \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}\Big[c^T\tilde{x} + Q(\tilde{x}, z)\Big] \\
&\leq \sup_{\mathbb{P}_z \in \mathcal{P}_z} \mathbb{E}_{\mathbb{P}_z}\Big[c^T\tilde{x} + \hat{Q}(\tilde{x}, z)\Big] + \|Q - \hat{Q}\|_\infty \\
&= \mathbb{E}_{\mathbb{P}_{\bar{z}}}\Big[c^T\tilde{x} + \hat{Q}(\tilde{x}, \bar{z}))\Big] + \|Q - \hat{Q}\|_\infty \qquad (5.36) \\
&\leq \mathbb{E}_{\mathbb{P}_{\bar{z}}}\Big[c^T\tilde{x} + Q(\tilde{x}, \bar{z})\Big] + 2\|Q - \hat{Q}\|_\infty \\
&= \tilde{\eta} + 2\|Q - \hat{Q}\|_\infty.
\end{aligned}$$

where the first inequality holds since $\tilde{x}$ is not necessarily optimal in model (5.10), and where we apply the result of Ben-Tal and Hochman (1972) in (5.36). ■

5.B Solution methods for the OR experiment of Section 5.5.1

In this appendix we discuss several solution methods for solving the OR scheduling problem.

5.B.1 SAA method

The main difficulty in solving the problem in (5.29) is to deal with the 3^N scenarios of the surgery durations $\bar{z}$. A well known approach in the SP literature to circumvent this difficulty is to use sampling to approximate $\bar{z}$ by a random vector having fewer scenarios. Thus, we sample N_s scenarios from the worst-case distribution $\mathbb{P}_{\bar{z}}$ to obtain the sample z^s, $s = 1, \ldots, N_s$. Then, letting $\theta_{i,s}$ denote the minutes of overtime in the i-th OR under scenario s, we can derive a large-scale deterministic equivalent MILP formulation:

$$\begin{aligned}
(DEF) \quad \min_{x,y,\theta} \quad & \sum_{i=1}^{N} c_f x_i + \frac{1}{N_s}\sum_{s=1}^{N_s}\sum_{i=1}^{N} c_v \theta_{i,s} \\
\text{s.t.} \quad & \theta_{i,s} \geq \sum_{j=1}^{n} z_j^s y_{ij} - T x_i, && i = 1, \ldots, N,\ s = 1, \ldots, N_s, \\
& (x, y) \in X,\ \ \theta_{i,s} \geq 0, && i = 1, \ldots, N,\ s = 1, \ldots, N_s.
\end{aligned}$$

This deterministic equivalent formulation contains $N + N^2$ binary variables, corresponding to x and y, and $N \times N_s$ continuous variables, corresponding to θ. In the numerical experiments in Section 5.5.1.2 we solve this MILP for $N_s = 1000$ and $N = 10$ or $N = 15$ using Gurobi. For these parameters the number of binary variables in the deterministic equivalent formulation is small so that the MILP can be solved

within reasonable time limits. Of course, using decomposition algorithms, such as for example an L-shaped algorithm, we may solve this model faster. However, we prefer to use the current method to show that good solutions may be obtained (for problems of reasonable size) using this straightforward, easy-to-implement algorithm.

5.B.2 Linear decision rules

Another way to deal with the 3^N scenarios is to use *linear decision rules* for the overtime costs θ_i. Instead of using the exact expression for θ_i, given in (5.27), we approximate θ_i by an affine function $\hat{\theta}_i$ of $\bar{z}$:

$$\hat{\theta}_i(z) = u_i + \sum_{j=1}^{N} v_{ij}\bar{z}_j. \tag{5.37}$$

Here, u_i and v_{ij} denote the coefficients of the linear decision rule $\hat{\theta}_i$. These coefficients are determined here-and-now, i.e. at the same time as the surgery-to-OR assignment (x, y). Hence, the first-stage decision variables in the resulting (approximating) optimization problem are $(x, y) \in X$ and the coefficients u and V, where $u = (u_1, \ldots, u_N)^\top$ and V is a matrix containing the elements v_{ij} for $i, j = 1, \ldots, N$. The approximating optimization problem using $\hat{\theta}_i$ instead of θ_i is given by

$$\begin{aligned}
(LDR) \quad \min_{x,y,u,V} \quad & \sum_{i=1}^{N} c_f x_i + c_v \mathbb{E}_{\mathbb{P}_{\bar{z}}}\Big[\sum_{i=1}^{N} \hat{\theta}_i(\bar{z})\Big] \\
\text{s.t.} \quad & \hat{\theta}_i(\bar{z}) \geq \sum_{j=1}^{N} \bar{z}_j y_{ij} - T x_i, && \bar{z} \in \mathcal{Z},\ i = 1, \ldots, N, \quad (5.38)\\
& \hat{\theta}_i(\bar{z}) \geq 0, && \bar{z} \in \mathcal{Z},\ i = 1, \ldots, N, \quad (5.39)\\
& (x, y) \in X.
\end{aligned}$$

Here, constraints (5.38) and (5.39) make sure that for every $i = 1, \ldots, N$ the approximation $\hat{\theta}_i(\bar{z})$ for the overtime costs is non-negative and at least as large as the actual overtime costs $\theta_i(\bar{z})$ for all 3^N possible realizations of $\bar{z}$. Moreover, using the linear decision rule in (5.37), the expected overtime costs in the objective of (LDR) become

$$\begin{aligned}
c_v \mathbb{E}_{\mathbb{P}_{\bar{z}}}\Big[\sum_{i=1}^{N} \hat{\theta}_i(\bar{z})\Big] &= c_v \mathbb{E}_{\mathbb{P}_{\bar{z}}}\Big[\sum_{i=1}^{N} \Big(u_i + \sum_{j=1}^{N} v_{ij}\bar{z}_j\Big)\Big] \\
&= c_v \sum_{i=1}^{N} \Big(u_i + \sum_{j=1}^{N} v_{ij}\mu_j\Big) \\
&= c_v \sum_{i=1}^{N} u_i + c_v \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}\mu_j.
\end{aligned}$$

The optimization problem (LDR) can thus be rewritten as

$$
\begin{aligned}
(LDR) \quad \min_{x,y,u,V} \quad & c_f \sum_{i=1}^{N} x_i + c_v \sum_{i=1}^{N} u_i + c_v \sum_{i=1}^{N} \sum_{j=1}^{N} v_{ij}\mu_j \\
\text{s.t.} \quad & u_i + \sum_{j=1}^{N} v_{ij}\bar{z}_j \geq \sum_{j=1}^{N} \bar{z}_j y_{ij} - T x_i, && \bar{z} \in \mathcal{Z},\ i = 1, \ldots, N, \qquad (5.40) \\
& u_i + \sum_{j=1}^{N} v_{ij}\bar{z}_j \geq 0, && \bar{z} \in \mathcal{Z},\ i = 1, \ldots, N, \qquad (5.41) \\
& (x, y) \in X.
\end{aligned}
$$

Observe that this is a MILP with only a small number of decision variables, but with exponentially many constraints, since (5.40) and (5.41) are defined for every $\bar{z} \in \mathcal{Z}$, and $\bar{z}$ is a discrete random vector with 3^N realizations. However, since the convex hull of Z is a box uncertainty set, we can use standard techniques from Robust Optimization (Ben-Tal et al. 2009) to obtain the robust counterpart of this problem.

5.B.3 Adjusted version of the L-shaped algorithm

The L-shaped algorithm solves the optimization problem in (5.29) exactly. In this section we discuss our tailored implementation of the L-shaped algorithm, see e.g. van Slyke and Wets (1969). In this algorithm we approximate the expected overtime costs

$$
Q(x, y) = c_v \mathbb{E}_{\mathbb{P}_{\bar{z}}}\Big[\sum_{i=1}^{N} \Big(\sum_{j=1}^{N} \bar{z}_j y_{ij} - T x_i \Big)^+ \Big], \quad (x, y) \in X,
$$

by an artificial decision variable $\eta \geq 0$. Using optimality cuts of the form

$$
\eta \geq \pi^k x + \lambda^k y + \beta^k, \quad k = 1, \ldots, K,
$$

where π^k and λ^k are row vectors and β^k is a constant, the variable η will represent a lower bound of $Q(x, y)$.

At every iteration k, we solve the master problem

$$
\begin{aligned}
(\mathrm{MP}_k) \quad \min_{x,y,\eta} \quad & c_f \sum_{i=1}^{N} x_i + \eta \\
\text{s.t.} \quad & \eta \geq \pi^l x + \lambda^l y + \beta^l, && l = 1, \ldots, k-1, \\
& (x, y) \in X, \quad \eta \geq 0,
\end{aligned}
$$

which includes all optimality cuts of previous iterations, to obtain the current solution (x^k, y^k, η^k). For this current solution, we evaluate $Q(x^k, y^k)$. Note that η^k is a lower bound of $Q(x^k, y^k)$ and that if $Q(x^k, y^k) = \eta^k$, then (x^k, y^k) is the optimal solution to (5.29). In general, we stop the algorithm if $Q(x^k, y^k) - \eta^k < \epsilon$ for some small number ϵ. If this optimality criterion does not hold, then we derive an optimality cut $\eta \geq \pi^k x + \lambda^k y + \beta^k$ and solve the master problem MP_{k+1}.

The challenge in this algorithm is to evaluate $Q(x^k, y^k)$ and to derive an optimality cut $\eta \geq \pi^k x + \lambda^k y + \beta^k$, dealing with the 3^N scenarios of the random vector $\bar{z}$. We only discuss why $Q(x^k, y^k)$ may be evaluated fast, despite this exponential number of scenarios. For similar reasons we may derive optimality cuts in an efficient way.

5.B.3.1 Evaluation of $Q(x^k, y^k)$

In this section we discuss the evaluation of $Q(x^k, y^k)$. For convenience we drop the index k and use (x, y) to refer to the current solution. We speed up the evaluation of $Q(x, y)$ in two different ways. First, by using the simple recourse structure of the problem, and second by using an efficient data structure for the scenarios.

Since the operating room scheduling problem can be modeled as a simple integer recourse model with uncertainty in the technology matrix, we can use the results of Klein Haneveld and van der Vlerk (2006) to evaluate $Q(x, y)$ faster. Similar as in Section 5.2 this is possible since we are not dealing with the joint distribution of $\bar{z}$ but with several marginal distributions of total surgery durations in operating rooms. To show this more formally, we first introduce additional notation.

Let $J_i = \{j : y_{ij} = 1\}$ denote the set of surgeries carried out in the i-th OR with $N_i := |J_i|$ denoting the number of surgeries. Moreover, let ζ^i denote the subvector of $\bar{z}$ containing these N_i surgeries. That is, ζ^i contains the j-th component of $\bar{z}$ if and only if $j \in J_i$. Then, by separability of the expected overtime costs,

$$\begin{aligned} Q(x, y) &= \mathbb{E}_{\mathbb{P}_{\omega^*}}\Big[\sum_{i=1}^{N} Q_i(x, y)\Big] \\ &= \sum_{i=1}^{N} \mathbb{E}_{\mathbb{P}_{\zeta^i}}\Big[\Big(\sum_{j=1}^{N} \zeta_j^i y_{ij} - T x_i\Big)^+\Big]. \end{aligned}$$

Since the marginal distribution of ζ^i has 3^{N_i} realizations, it follows immediately that it suffices to compute overtime costs for $\sum_{i=1}^{N} 3^{N_i}$ 'scenarios'. For example, if $N_i = 1$ for all $i = 1, \ldots, N$, i.e. if in every OR only a single surgery is carried out, then this number reduces to $3N$, whereas if $N_i = N$, then it equals 3^N.

The second way to speed up computations is based on the following two special cases, which we consider for the i-th OR only. If overtime costs are zero for every scenario,

then $Q_i(x, y) = 0$, and if overtime costs are positive for all scenarios, then

$$Q_i(x, y) = \mathbb{E}_{\mathbb{P}_{\zeta^i}}\Big[\Big(\sum_{j=1}^{N} \zeta_j^i y_{ij} - Tx_i\Big)^+\Big] = \sum_{j \in J_i} \Big(\mu_j y_{ij} - Tx_i\Big).$$

The first case hold if $\sum_{j \in J_i} b_j \leq T$ and the second case if $\sum_{j \in J_i} a_j > T$.

The main idea for a general approach, exploiting these special cases, is to iteratively condition on surgery durations ζ_l^i until one of the two special cases applies, i.e. until overtime costs corresponding to all scenarios under consideration are either all zero or all positive. In this way we do not necessarily have to compute all overtime costs for each scenario.

An alternative interpretation of the same idea is to assume that the random vector ζ^i is ordered chronologically, meaning that the first component of ζ^i corresponds to the surgery that is carried out first and the last component of ζ^i corresponds to the surgery that is carried out last. Thus, surgery durations are revealed gradually over time, and we may represent this process by a scenario tree, see e.g. Figure 5.1. This scenario tree represents all possible realizations of surgery durations at every stage $l = 0, \ldots, N_i$, where stage l corresponds to the situation where the first l surgery durations have been observed (i.e. the first l surgeries have been carried out). For example, at the root node in stage 0, no surgery durations are observed yet, whereas in the leave nodes at stage N_i all surgery durations are completely specified.

We iteratively construct the scenario tree, keeping track of the probability $p(n)$ of reaching each node n, and the total surgery duration $D(n)$ of all surgeries carried out before reaching node n. For every such node n at stage l, we compute whether

$$(i) \quad D(n) + \sum_{j=l+1}^{N_i} b_j \leq T,$$

and

$$(ii) \quad D(n) + \sum_{j=l+1}^{N_i} a_j > T.$$

If (i) holds, then the average overtime costs $q(n)$ over all scenarios corresponding to subleaves of node n equal zero. If (ii) holds, then

$$q(n) = D(n) + \sum_{j=l+1}^{N_i} \mu_j,$$

and if (i) and (ii) both do not hold, then we expand the scenario tree, creating three subnodes of n at stage $l+1$, by conditioning on the three possible realizations of the $(l+1)$-th surgery duration. We repeat this process until all nodes are evaluated. In practice much fewer evaluations than 3^{N_i} will be required, as shown in Table 5.4 in the main part of the paper.

5.C Error bounds for the integer experiment of Section 5.5.2

5.C.1 General description

In this appendix we discuss how we obtain the error bounds in Tables 5.5 and 5.6 of Section 5.5.2. This error bound is derived by combining the Multiple Replications Procedure (MRP) of Bayraksan and Morton (2009) and the total variation error bounds discussed in Section 5.3. To our knowledge, this is the first attempt to combine these two approaches. Moreover, it may be interesting to apply this error bound to other applications involving integer decision variables and uncertain random parameters.

5.C.2 A straightforward error bound

Before we derive this error bound, we first discuss why direct application of the error bound of Section 5.3 is not sufficient to obtain a tight bound.

Let $(\hat{x}, \hat{y})$ denote the optimal surgery-to-OR assignment in (5.30) and let η^* denote the optimal objective value of the original problem. Then, by Theorem 5.1 in Section 5.3 and defining

$$G(\hat{x}, \hat{y}) := \sum_{i=1}^{N} c_f \hat{x}_i + \sup_{\mathbb{P}_\epsilon \in \mathcal{P}_\epsilon} \mathbb{E}_{\mathbb{P}_\epsilon} Q(\hat{x}, \hat{y}, \epsilon), \tag{5.42}$$

we have

$$G(\hat{x}, \hat{y}) - \eta^* \leq 2\|Q - \hat{Q}\|_\infty := 2 \sup_{x,y,\epsilon} |Q(x, y, \epsilon) - \hat{Q}(x, y, \epsilon)|.$$

We can obtain an upper bound on $|Q(x, y, \epsilon) - \hat{Q}(x, y, \epsilon)|$ by straightforward application of the total variation error bounds derived in Romeijnders et al. (2016b). However, this bound depends significantly on the surgery-to-OR assignment (x, y). For example, if every surgery is carried out in a separate OR then the bound reduces to

$$|Q(x, y, \epsilon) - \hat{Q}(x, y, \epsilon)| \leq \frac{1}{2} \sum_{j=1}^{N} 60 c_v h(60|\Delta| f_j), \tag{5.43}$$

where f_j is the marginal density of the random surgery duration z_j in minutes, and $h(t) = t/8$ if $t \leq 4$ and $h(t) = 1 - 2/t$, otherwise. The value of 60 is present in the error bound since overtime wages are paid in full hours (of 60 minutes). In contrast, if all surgeries are carried out in a single OR, then the error bound reduces to

$$|Q(x, y, \epsilon) - \hat{Q}(x, y, \epsilon)| \leq 30 c_v h(60|\Delta| \bar{g}), \tag{5.44}$$

where $\bar{g}$ is the marginal density of *the sum* of all surgery durations.

For the numerical experiments in Section 5.5.2, the bound in (5.43) is much larger than the bound in (5.44). Both, however, are by definition larger than the upper bound on $G(\hat{x},\hat{y})-\eta^*$. This implies that the error bound $2\|Q-\hat{Q}\|_\infty$ may too large for practical purposes. However, at the same time we do not expect such an extreme surgery-to-OR assignment (x,y) as in (5.43) to be optimal. This is relevant, since for computing a valid error bound we only require the difference between Q and $\hat{Q}$ in the approximating solution $(\hat{x},\hat{y})$ and the optimal solution (x^*,y^*). The problem, however, is that we do not know the optimal solution (x^*,y^*). Thus, although we do know the approximating solution $(\hat{x},\hat{y})$, to obtain a valid upper bound we need to take into account the worst-case surgery-to-OR assignment.

5.C.3 Multiple Replications Procedure

To avoid this problem we will apply the Multiple Replications Procedure (MRP) described in e.g. Bayraksan and Morton (2009). This method cannot be readily applied since it requires us to determine the worst-case probability distribution in the integer model, but combined with the total variation error bound it will yield a much tighter (probabilistic) bound. Below we describe the main outline of the approach.

To assess the quality of the approximating solution we will use (an adjusted version) of the Multiple Replications Procedure (MRP) described in e.g. Bayraksan and Morton (2009). The goal is to evaluate $G(\hat{x},\hat{y})-\eta^*$, where G is defined in (5.42). Since Q is not convex in ϵ we cannot use the results of this paper to determine the worst-case distribution of ϵ. That is why we approximate Q by $\hat{Q}$ and obtain

$$\begin{aligned}
G(\hat{x},\hat{y}) &= \sum_{i=1}^{N} c_f\hat{x}_i + \sup_{\mathbb{P}_\epsilon\in\mathcal{P}_\epsilon}\left\{\mathbb{E}_{\mathbb{P}_\epsilon}\left[\hat{Q}(\hat{x},\hat{y},\epsilon)+\left(Q(\hat{x},\hat{y},\epsilon)-\hat{Q}(\hat{x},\hat{y},\epsilon)\right)\right]\right\}\\
&\leq \sum_{i=1}^{N} c_f\hat{x}_i + \sup_{\mathbb{P}_\epsilon\in\mathcal{P}_\epsilon}\mathbb{E}_{\mathbb{P}_\epsilon}\hat{Q}(\hat{x},\hat{y},\epsilon) + \sup_{\mathbb{P}_\epsilon\in\mathcal{P}_\epsilon}\mathbb{E}_{\mathbb{P}_\epsilon}\left[Q(\hat{x},\hat{y},\epsilon)-\hat{Q}(\hat{x},\hat{y},\epsilon)\right]\\
&= \hat{G}(\hat{x},\hat{y}) + \sup_{\mathbb{P}_\epsilon\in\mathcal{P}_\epsilon}\mathbb{E}_{\mathbb{P}_\epsilon}\left[Q(\hat{x},\hat{y},\epsilon)-\hat{Q}(\hat{x},\hat{y},\epsilon)\right],
\end{aligned}$$

where $\hat{G}$ equals G with Q replaced by $\hat{Q}$. Since $\hat{Q}$ is convex in ϵ it follows that $\mathbb{P}_{\bar{\epsilon}}$ is the worst-case distribution of ϵ in $\hat{G}$. This implies that $\hat{G}(\hat{x},\hat{y})$ does not contain any optimization problem and that it can easily be estimated using (Monte Carlo) sampling.

To eliminate the supremization of $\mathbb{P}_\epsilon$ over $\mathcal{P}_{(\mu,d)}$ in η^* we assume that $\mathbb{P}_{\bar{\epsilon}}$ is the worst-case distribution, so that we obtain a lower bound $\tilde{\eta}$ for η^*, see Theorem 5.1 in Section 5.3. To obtain $\tilde{\eta}$ we still have to minimize over all feasible surgery-to-OR assignments $(x,y)\in X$. However, the MRP is able to deal with such problems.

Combining both results we obtain an upper bound on the optimality gap:

$$G(\hat{x},\hat{y}) - \eta^* \leq \hat{G}(\hat{x},\hat{y}) - \tilde{\eta} + \sup_{\mathbb{P}_\epsilon \in \mathcal{P}_\epsilon} \mathbb{E}_{\mathbb{P}_\epsilon}\Big[Q(\hat{x},\hat{y},\epsilon) - \hat{Q}(\hat{x},\hat{y},\epsilon)\Big].$$

We will use the MRP to bound $\hat{G}(\hat{x},\hat{y}) - \tilde{\eta}$ and we use a total variation error bound for the last term. This bound may be much tighter than the previous total variation bound, since we only have to compute the difference between Q and $\hat{Q}$ for a *fixed* surgery-to-OR assignment $(\hat{x},\hat{y})$.

5.C.4 Total variation bounds

In this section we consider the total variation error bounds mentioned in the previous sections. Suppose that a feasible surgery-to-OR assignment (x,y) is given and assume for the moment that ϵ is fixed. We consider

$$Q(x,y,\epsilon) - \hat{Q}(x,y,\epsilon) = c_v \sum_{i=1}^{N} \psi_i(x,y,\epsilon),$$

where

$$\psi_i(x,y,\epsilon) = \mathbb{E}_{\mathbb{P}_z}\left[60\left[\Big(\sum_{j=1}^{N} z_j y_{ij} - (T+\epsilon_i)x_i\Big)/60\right]^+ - \Big(\sum_{j=1}^{N} z_j y_{ij} - (T+\epsilon_i-30)x_i\Big)^+\right].$$

Obviously, if $x_i = 0$, then $y_{ij} = 0$ for all $j = 1,\ldots,N$, so that $\psi_i(x,y,\epsilon) = 0$. If not, then let J_i denote the set of surgeries that are carried out in OR i, i.e. $j \in J_i$ if and only if $y_{ij} = 1$. Then, define ξ_i as the total surgery duration in the i-th OR in minutes:

$$\xi_i = \sum_{j=1}^{N} z_j y_{ij} = \sum_{j\in J_i} \omega_j.$$

Using this definition, and defining $s_i = (T+\epsilon_i)x_i = T + \epsilon_i$, the expression for ψ_i reduces to

$$\psi_i(x,y,\epsilon) = \mathbb{E}_{\mathbb{P}_{\xi_i}}\left[60\left[\frac{\xi_i - s_i}{60}\right]^+ - \Big(\xi_i - s_i + 30\Big)^+\right].$$

Applying the total variation error bound yields

$$\psi_i(x,y,\epsilon) \leq 30h(60|\Delta|g_i), \tag{5.45}$$

where $|\Delta|g_i$ is the total variation of the probability density function g_i of the total surgery duration ξ_i in the i-th OR. Thus,

$$Q(x,y,\epsilon) - \hat{Q}(x,y,\epsilon) \leq \frac{1}{2} c_v \sum_{i=1}^{N} 60h(|\Delta|g_i)x_i. \tag{5.46}$$

Here, we add x_i in the expression to ensure that $\psi_i(x, y, \epsilon) = 0$ if $x_i = 0$. Observe that if each surgery is carried out in a separate OR, that in this case $g_i = f_i$ for every $i = 1, \ldots, N$ and we obtain the bound given in (5.43). On the other hand, if all surgeries are carried out in a single OR, then we obtain the bound in (5.44). Since the bound in (5.46) holds independent of the value of ϵ, we conclude that

$$\begin{aligned}\sup_{\mathbb{P}_\epsilon \in \mathcal{P}_\epsilon} \mathbb{E}_{\mathbb{P}_\epsilon}\Big[Q(\hat{x}, \hat{y}, \epsilon) - \hat{Q}(\hat{x}, \hat{y}, \epsilon)\Big] &\leq \sup_{\mathbb{P}_\epsilon \in \mathcal{P}_\epsilon} \mathbb{E}_{\mathbb{P}_\epsilon}\Big[\frac{1}{2} c_v \sum_{i=1}^{N} 60h(|\Delta| g_i)\hat{x}_i\Big] \\ &= \frac{1}{2} c_v \sum_{i=1}^{N} 60h(|\Delta| g_i)\hat{x}_i. \end{aligned} \tag{5.47}$$

This error bound may be much smaller than the one described in (5.43). For one, since $\hat{x}_i$ may be zero for many ORs. In addition, g_i is the pdf of the sum of several independent random variables, and increasing the number of surgeries in the i-th OR will decrease the total variation of g_i.

5.C.4.1 Tighter total variation error bounds

The error bounds in Tables 5.5 and 5.6 are still even tighter than the one in (5.47). Surprisingly, we derive these bounds by applying the result of Ben-Tal and Hochman (1972) once more.

First, consider again

$$\psi_i(x, y, \epsilon) = \mathbb{E}_{\xi_i}\left[60 \left\lceil \frac{\xi_i - s_i}{60} \right\rceil^+ - \Big(\xi_i - s_i + 30\Big)^+\right],$$

where $s_i = T + \epsilon_i$ and observe that its underlying value function equals zero if $\xi_i \leq s_i - 30$. For this reason, omitting the technical details, we show that for fixed $s_i = T + \epsilon_i \in \mathbb{R}$,

$$\begin{aligned}\psi_i(x, y, \epsilon) &\leq 30h\Big(60|\Delta| g_i\Big([s_i - 30, +\infty)\Big)\Big) \\ &= 30h\Big(60|\Delta| g_i\Big([T + \epsilon_i - 30, +\infty)\Big)\Big),\end{aligned}$$

where $g_i([T + \epsilon_i - 30, +\infty))$ denotes the total variation of g_i on the interval $[T + \epsilon_i - 30, +\infty)$. This bound is tighter than the one in (5.45) and is attained if $\epsilon_i \to -\infty$. Since the bound is non-increasing in ϵ_i, we may conclude that for every $\epsilon_i \in [a_i, b_i]$,

$$\psi(x, y, \epsilon) \leq 30h\Big(60|\Delta| g_i\Big([T + a_i - 30, +\infty)\Big)\Big),$$

and thus

$$\sup_{\mathbb{P}_\epsilon \in \mathcal{P}_\epsilon} \mathbb{E}_{\mathbb{P}_\epsilon}\Big[Q(\hat{x}, \hat{y}, \epsilon) - \hat{Q}(\hat{x}, \hat{y}, \epsilon)\Big] \leq \frac{1}{2} \sum_{i=1}^{N} 60 c_v h\Big(60|\Delta| g_i\Big([T + a_i - 30, +\infty)\Big)\Big)\hat{x}_i.$$

However, we may obtain a tighter bound if the bound on $\psi(x, y, \epsilon)$ is convex in ϵ_i for $\epsilon_i \in [a_i, b_i]$ since it allows us to apply the result of Ben-Tal and Hochman (1972) in a surprising way. For example, if for all opened ORs i, the bounds are convex for $\epsilon_i \in [a_i, b_i]$, then

$$\sup_{\mathbb{P}_\epsilon \in \mathcal{P}_\epsilon} \mathbb{E}_{\mathbb{P}_\epsilon}\Big[Q(\hat{x}, \hat{y}, \epsilon) - \hat{Q}(\hat{x}, \hat{y}, \epsilon)\Big] \leq \frac{1}{2}\mathbb{E}_{\mathbb{P}_{\bar{\epsilon}}}\Big[\sum_{i=1}^{N} 60 c_v h\Big(60|\Delta| g_i\Big([T + \bar{\epsilon}_i - 30, +\infty)\Big)\Big)\hat{x}_i\Big]. \tag{5.48}$$

Of course, the bound $h(60|\Delta|g_i([T+\epsilon_i-30, +\infty))$ is in general not convex, but it may be in special cases. Notice, for example, that h is linear on $[0, 4]$ so that the bound is convex if $|\Delta|g_i([t + \epsilon_i - 30, +\infty))$ is convex in $\epsilon_i \in [a_i, b_i]$ and this total variation is small enough. In our numerical experiments, g_i is the pdf of the sum of several independent lognormal random variables, so that by the Central Limit Theorem, it is approximately normally distributed. Since a normal density function has a convex decreasing right tail it may satisfy the requirements. In our numerical experiments we check numerically for every opened OR i whether convexity holds; if not then we replace ϵ_i^* by a_i in the error bound of (5.48).

5.D Inventory experiment

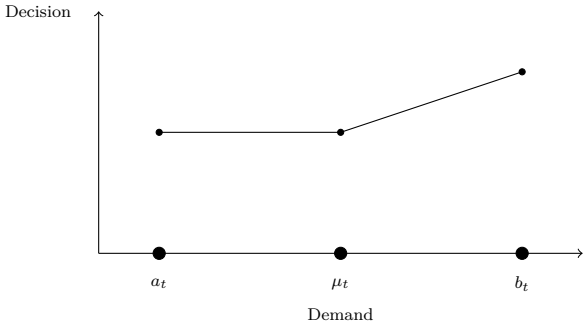
5.D.1 Decisions for uncertainty realizations outside the finite worst-case support

In this Appendix, we provide a detailed procedure to obtain a sequence of feasible decisions for arbitrary outcome of uncertainty $\hat{z}$, based on the solution to worst-case expectation version of problem (5.7) with 3^T points in the support. The idea is to use the solutions $\overline{x}_1, \overline{x}_2, ...$ to (5.7) and the convexity of the feasible set of (5.7) to construct a feasible sequence of decisions for all stages. In this setting, the later-stage decisions become piecewise-affine functions of the observed uncertainties.

The way we accomplish this is the following. Each realization of $\hat{z}$ is a convex combination of some of the elements z_t of the discrete worst-case distribution support. For a given component $\hat{z}_t \in \mathbb{R}$ we define that (i) if $\hat{z}_t \in [a_t, \mu_t]$ then $\hat{z}_t$ is formulated as a convex combination of a_t and μ_t, (ii) if $\hat{z}_t \in [\mu_t, b_t]$ then $\hat{z}_t$ is formulated as a convex combination of μ_t and b_t. Then, for each realization $\hat{z}_t$ we obtain a unique set of coefficients of the convex combination. A way to implement feasible decisions corresponding to an arbitrary realization $\hat{z}_t$ is to use the same convex combination coefficients with respect to decision vectors corresponding to the uncertainty realizations in the support of the worst-case distribution of z, illustrated in Figure 5.3.

The following example illustrates this idea mathematically.

Figure 5.3 – Piecewise linear decision rules. Having the defined the decisions for a_t, μ_t, and b_t, the decision for points outside $\{a_t, \mu_t, b_t\}$ are convex combinations of the decisions for a_t, μ_t or μ_t, b_t.



Example 5.2 *Imagine a problem instance where $T = 2$ and $z \in \mathbb{R}$ belongs to a (μ, d) ambiguity set with $a = 0$, $\mu = 1$, and $b = 2$ and the realized uncertainty is $\hat{z} = 0.5$. Then we can see that $\hat{z} \in [a_1, \mu_1]$. So, we have that $\hat{z}$ is a convex combination of a_1, and μ_1. Indeed, take*

$$\hat{z} = \lambda a_1 + (1 - \lambda)\mu_1.$$

Since each of the points a, μ has its own stage 2 decision, let us denote them as $x_2(a)$ and $x_2(\mu)$:

$$x_2(\hat{z}) = \lambda x_2(a) + (1 - \lambda)x_2(\mu).$$

Note that such a policy is feasible for $\hat{z}$ because:

$$\begin{aligned} & A_{2,1}(a)x_1 + A_{2,2}x_2(a) = b_2(a), \quad A_{2,1}(\mu)x_1 + A_{2,2}x_2(\mu) = b_2(\mu) \quad \Rightarrow \\ \Rightarrow \quad & A_{2,1}(\lambda a_1 + (1-\lambda)\mu_1)x_1 + A_{2,2}x_2(\lambda a_1 + (1-\lambda)\mu_1) = b_2(\lambda a_1 + (1-\lambda)\mu_1) \end{aligned}$$

by linearity of $A_{2,1}(\cdot)$ and $b_2(\cdot)$.

More formally, the decisions are implemented as:

$$x(\hat{z}) = \sum_{\kappa \in \{1,2\}^{n_z}} \lambda(\kappa) x(z(\kappa)),$$

where $x(z(\kappa))$ is the sequence of decisions related to the demand trajectory $z(\kappa)$ in the worst-case support, defined as:

$$z(\kappa) = \begin{bmatrix} z_1(\kappa_1) \\ \vdots \\ z_{T-1}(\kappa_{T-1}) \end{bmatrix}, \quad \kappa = \begin{bmatrix} \kappa_1 \\ \vdots \\ \kappa_{T-1} \end{bmatrix} \quad z(\kappa_t) = \begin{bmatrix} z_{t1}(\kappa_{t1}) \\ \vdots \\ z_{tn_{z,t}}(\kappa_{t,n_{z,t}}) \end{bmatrix}, \quad \kappa_t = \begin{bmatrix} \kappa_{t1} \\ \vdots \\ \kappa_{tn_{z,t}} \end{bmatrix}$$

and

$$z_{tj}(1) = \begin{cases} a_{tj} & \text{if } \hat{z}_{tj} \leq \mu_{tj} \\ \mu_{tj} & \text{otherwise,} \end{cases}$$

$$z_{tj}(2) = \begin{cases} \mu_{tj} & \text{if } \hat{z}_{tj} \leq \mu_{tj} \\ b_{tj} & \text{otherwise.} \end{cases}$$

and $\lambda(\kappa)$ are unique coefficients such that:

$$\hat{z} = \sum_{\kappa \in \{1,2\}^{n_z}} \lambda(\kappa) z(\kappa),$$

where

$$\lambda(\kappa) = \prod_{t=1}^{T-1} \lambda_t(\kappa_t), \quad \lambda_t(\kappa_t) = \prod_{j=1}^{n_{z,t}} \lambda_{tj}(\kappa_{tj}), \quad \lambda_{tj}(1) = \begin{cases} \frac{\hat{z}_{tj} - a_{tj}}{\mu_{tj} - a_{tj}} & \text{if } \hat{z}_{tj} \leq \mu_{tj} \\ \frac{\hat{z}_{tj} - \mu_{tj}}{b_{tj} - \mu_{tj}} & \text{otherwise,} \end{cases}$$

and $\lambda_{tj}(2) = 1 - \lambda_{tj}(1)$. In this way the resulting decision always satisfies the problem's constraints for a given realization of the uncertain parameter.

5.D.2 Reoptimization to compute the intervals of Section 5.5.3.2

We explain the computation of the ends of the intervals using reoptimization on the example of computing the lower bound for the PL-WCE solutions. The question corresponding to computation of the lower end of the interval is: what is the expected total cost if the true demand distribution is the best-case distribution, but the decision maker assumes all the time that it is the worst-case distribution?

To answer it, for each of the 2^T best case demand trajectories we compute the ordering decisions in a reoptimizing fashion. That is, at time 1 decisions are determined such that the worst-case expectation is minimized. The corresponding time 1 decision is implemented. At time 2 it turns out that the demand in time 1 was one of the demands belonging to the best-case support $\{\mu_1 - d_1/2(1-\beta_1), \mu_1 + d_1/2\beta_1\}$ of z_1. In this situation, decisions for stages 2-6 are constructed (thus, a new optimization problem is solved) that minimize again the worst-case expectation and the corresponding time 2 decision is implemented. At time 3 it turns out that the demand at time 2 belonged to the support $\{\mu_2 - d_2/2(1-\beta_2), \mu_2 + d_2/2\beta_2\}$ of the best-case distribution and so on. In the end, for each of the 2^T demand trajectories in the support of the best-case distribution, a corresponding decision trajectory is obtained and the objective function values for each of the best-case trajectories are weighted with the corresponding best-case probability.

5.D.3 Simulating the (μ, d) sample

The (μ, d) sample in the inventory experiment is constructed as follows. First, a discretized distribution $\hat{\mathbb{P}} \in \mathcal{P}_z$ is sampled using the hit-and-run method (Smith 1984). The hit-and-run method is implemented as follows. For the $[0, 1]$ interval (from which the demands on the relevant support intervals are sampled using the inverse transform) we construct a grid of 51 equidistant points. For a fixed (μ, d) the set of probability masses for the points of the grid satisfying the μ and d values constitutes a polytope. We sample 10 probability distributions uniformly from this polyhedron by iteratively choosing a random direction and sampling uniformly a point on the segment of the line along this direction belonging to the polytope. Then, we sample the demand in each period in two steps, by sampling first one of the distributions, and then by sampling a point in the $[0, 1]$ interval with the given distribution.

Bibliography

S. Ahmed, M. Tawarmalani, and N. V. Sahinidis. A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming*, 100:355–377, 2004.

G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs via sampling. In *Tutorials in Operations Research*, pages 102–122. INFORMS, 2009.

A. Ben-Tal and E. Hochman. More bounds on the expectation of a convex function of a random variable. *Journal of Applied Probability*, 9:803–812, 1972.

A. Ben-Tal and E. Hochman. Stochastic programs with incomplete information. *Operations Research*, 24(2):336–347, 1976.

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

A. Ben-Tal, B. Golany, A. Nemirovski, and J.-Ph. 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 University Press, 2009.

A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 1-2:265–299, 2015.

D. Bertsimas and A. Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, 63(3):610–627, 2015.

D. Bertsimas, D. A. Iancu, and P. A. Parrilo. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12): 2809–2824, 2011.

D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2013.

J. R. Birge. The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming*, 24(3):314âĂŞ325, 1982.

J. R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33:989–1007, 1985.

J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

C. C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24:37–45, 1999.

X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6)(6):1469–1482, 2009.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

E. Delage, S. Arroyo, and Y. Ye. The value of distribution information in distributionally robust optimization. http://web.hec.ca/pages/erick.delage/ValueDistributionInfo.pdf, 2015.

B. T. Denton, A. J. Miller, H. J. Balasubramanian, and T. R. Huschka. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, 58: 802–816, 2010.

J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95: 493–511, 2003.

M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106:423–432, 2006.

N. C. P. Edirisinghe. *Stochastic Programming: The State of the Art In Honor of George B. Dantzig*, chapter Stochastic Programming Approximations Using Limited Moment Information, with Application to Asset Allocation, pages 97–138. Springer New York, 2011.

H. P. Edmundson. Bounds on the expectation of a convex function of a random variable. Technical Report Paper 982, The RAND Corporation, 1956.

A. V. Fiacco and J. Kyparisis. Convexity and concavity properties of the optimal value function in parametric nonlinear programming. *Journal of Optimization Theory and Applications*, 48(1):95–126, 1986.

S. J. Garstka and R. J.-B. Wets. On decision rules in stochastic programming. *Mathematical Programming*, 7(1):117–143, 1974.

S. Gul, B. T. Denton, J. W. Fowler, and T. Huschka. Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations Management*, 20:406–417, 2011.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage robust binary programming. *Operations Research*, 63(4):877–891, 2015.

H. Heitsch and W. Römisch. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24:187–206, 2003.

J. L. Higle and S. Sen. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16:650–669, 1991.

K. Isii. On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14(1):185–197, 1963.

J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 1906.

J. H. B. Kemperman. The general moment problem, a geometric approach. *Annals of Mathematical Statistics*, 39:93–122, 1968.

W. K. Klein Haneveld and M. H. van der Vlerk. Stochastic integer programming: General models and algorithms. *Annals of Operations Research*, 85:39–57, 1999.

W. K. Klein Haneveld and M. H. van der Vlerk. Integrated chance constraints: reduced forms and an algorithm. *Computational Management Science*, 3(4):245–269, 2006.

W. K. Klein Haneveld, L. Stougie, and M. H. van der Vlerk. Simple integer recourse models: convexity and convex approximations. *Mathematical Programming*, 108:435–473, 2006.

F. Knight. *Risk, uncertainty and profit.* New York, 1921.

G. Laporte and F. V. Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13:133–142, 1993.

F. V. Louveaux and M. H. van der Vlerk. Stochastic programming with simple integer recourse. *Mathematical Programming*, 61:301–325, 1993.

M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89:251–271, 2001.

K. Postek and D. den Hertog. Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing*, 28(3): 553–574, 2016.

K. Postek, A. Ben-Tal, D. den Hertog, and B. Melenberg. Exact robust counterparts of ambiguous stochastic constraints under mean and dispersion information. CentER Discussion Paper No. 2015-030, available at: http://www.optimization-online.org/DB_FILE/2015/06/4946.pdf, 2015.

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

R. T. Rockafellar and R. J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16:119–147, 1991.

W. Romeijnders, L. Stougie, and M. H. van der Vlerk. Approximation in two-stage stochastic integer programming. *Surveys in Operations Research and Management Science*, 19:17–33, 2014.

W. Romeijnders, M. H. van der Vlerk, and W. K. Klein Haneveld. Convex approximations of totally unimodular integer recourse models: A uniform error bound. *SIAM Journal on Optimization*, 25:130–158, 2015.

W. Romeijnders, R. Schultz, M. H. van der Vlerk, and W. K. Klein Haneveld. A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM Journal on Optimization*, 26:426–447, 2016a.

W. Romeijnders, M. H. van der Vlerk, and W. K. Klein Haneveld. Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Mathematical Programming*, 157:3–46, 2016b.

W. Römisch. Stability of stochastic programming problems. In A. Ruszczyński and A. Shapiro, editors, *Handbooks in Operations Research and Management Science*, volume 10: Stochastic Programming, pages 483–554. Elsevier, 2003.

W. Römisch and R. Schultz. Decomposition of a multi-stage stochastic program for power dispatch. *ZAMM-Zeitschrift für Angewandte Mathematik und Mechanik*, 76(3):29–32, 1996.

A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333, 1986.

H. Scarf. A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, and H. E. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, 1958.

R. Schultz. Stochastic programming with integer variables. *Mathematical Programming*, 97:285–309, 2003.

S. Sen. Algorithms for stochastic mixed-integer programming models. In K. Aardal, G. L. Nemhauser, and R. Weismantel, editors, *Handbooks in Operations Research and Management Science*, volume 12: Discrete Optimization, pages 515–558. Elsevier, 2005.

A. Shapiro. *On duality theory of conic linear problems*, pages 135–165. Springer US, 2001.

A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14(4):1237–1249, 2004.

A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.

R. L. Smith. Efficient monte-carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32:1296–1308, 1984.

M. H. van der Vlerk. Convex approximations for complete integer recourse models. *Mathematical Programming*, 99:297–310, 2004.

R. M. van Slyke and R. Wets. *L*-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17: 638–663, 1969.

J. Žáčková. On minimax solutions of stochastic linear programming problems. *Časopis Pro Pěstováni Matematiky*, 91:423–430, 1966.

R. J.-B. Wets. Solving stochastic programs with simple recourse. *Stochastics*, 10:219–242, 1983.

W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

V. Zverovich, C. I. Fábián, E. F. D. Ellison, and G. Mitra. A computational study of a solver system for processing two-stage stochastic LPs with enhanced Benders decomposition. *Mathematical Programming Computation*, 4:211–238, 2012.

CHAPTER 6

Adjustable robust strategies for flood protection

6.1 Introduction

Managing flood risks is of vital importance for the Netherlands, a country which is for two-thirds flood prone. The areas at risk are protected by a large flood protection system, with dike rings consisting of dikes, dams, dunes, high grounds, and other water defense structures. Each year, about 1 billion euro is spent to maintain and improve this system.

Improving the flood defenses is necessary when flood protection standards are no longer met. This can be caused by for example a general deterioration of the flood defense, climate change but also because of changing protection standards. In the recent years, and within the context of the Dutch Delta Program, the Netherlands has been in the process of revising all its legal flood protection standards (Van Alphen 2015), since part of the old standards was still based on the advice of the first Delta Commission, installed after the last great flood disaster in the Netherlands of 1953 (Van Dantzig 1956).

In the Delta program, a novel, optimizing cost-benefit analysis (CBA) using operations research techniques was used to derive economically efficient ('optimal') flood protection standards, in which the discounted total cost of expected long-run investment and remaining flood damages was minimized (Brekelmans et al. 2012, Kind 2014, Eijgenraam et al. 2014, 2016). This CBA assumed dike reinforcements as flood protection measure and determined the optimal timing and size of the reinforcements, from which subsequently optimal flood protection standards were derived. It turned out that although the optimal timing and size of the reinforcements were dependent on the climate scenario assumed, the optimal standards were not. The optimal standards, however, are very sensitive to assumptions on e.g. the economic scenario, discount rate, investment costs and flood damages (Gauderis et al. 2013).

The updated standards, accepted by the Dutch government on July 5, 2016, are based on considerations over equity (a maximum tolerable probability for all individuals to lose his/her life because of flooding), efficiency (CBA) and social disruption (Van

der Most et al. 2014). After the new flood protection standards become legally binding from 2017 and onwards, regional delta programs have the task of developing more detailed flood risk management strategies to meet those new flood protection standards. Those strategies are not necessarily restricted to dike reinforcements, but other measures are also considered.

In the lower reaches of the Rhine and Meuse rivers, the regional Delta Program *Rhine Estuary - Drechtsteden* studies alternatives to dike reinforcements, such as water storage, channel deepening, storm surge barriers, and room for the river measures - alternatives which lower (design) water levels and hence reduce the need to reinforce and heighten the dikes (Jeuken et al. 2013). To aid the regional program in developing and evaluating combined strategies, including proposal for timing of the investments, a 'planning kit' was developed (Kind et al. 2014). It turned out there was a large number of strategies possible and it was practically infeasible to find an optimal strategy by trial and error, even if uncertainties in climate scenarios were not considered.

There are two important challenges remaining in designing the optimal flood risk management strategies. First, it is important to incorporate climate change related uncertainties in the analysis explicitly. In the previous approaches, sensitivity analysis was used to show robustness, that is, the performance of the given solution was tested against different realizations of the uncertainty. Such an approach may (i) take a lot of computation time and for each scenario requires a detailed specification of the parameter evolution and (ii) may be infeasible already at the implementation stage - because parameter values different from the assumed ones are revealed between the decision-making moment and the implementation moment. This raises the challenge of *robustness to parameter uncertainty*. Second, it is important that the solutions are *adjustable* to the revealed values of uncertainties as opposed to *static* solutions. The static solutions, where the later-stage decisions do not adapt to the true outcome of, say, sea level rise speed in the first 30 years, may prove to be over-conservative and expensive. This raises the challenge of *adjustability to revealed uncertainties*.

In this paper, we develop a mathematical optimization model to determine optimal adaptation strategies that addresses both of the challenges above. This model is general and applicable in any area, and in our experiment we apply it to the Rhine Estuary area, where it improves upon the planning kit of (Kind et al. 2014). The issue of climate change related uncertainties is addressed by using Robust Optimization (RO, see Ben-Tal et al. (2009)). In this approach, instead of specifying a precise climate change trajectory, an uncertainty set of 'possible outcomes' of the unknown parameters is specified. Then, the problem is solved in such a way that the constraints (requirements) are satisfied by the decisions for every outcome of uncertainty within the uncertainty set.

The strict requirement that the decisions are feasible for all allowed outcomes of uncertainty makes RO the preferred methodology for safety-related optimization problem such as flood prevention. For an introduction and overview of techniques used in RO, we refer the reader to the work by Ben-Tal et al. (2009), Bertsimas et al. (2011a), Gabrel et al. (2014) and references therein.

In order to allow the later-stage decisions adapt to the revealed uncertainties from earlier stage, we resort to an extension of RO - the Adjustable Robust Optimization (ARO). There, apart from requiring the constraints to hold for all outcomes of the uncertainty, the later stage decisions are formulated as functions of the uncertainties revealed before they are implemented, and the way these decisions adapt ('shape of the reaction') is also optimized.

ARO was initially developed to solve problems with continuous decision variables in Ben-Tal et al. (2004) where the concept of using affinely adjustable decision rules was introduced, extended later by Chen et al. (2012), Chen and Zhang (2012), Ben-Tal et al. (2009), and Bertsimas et al. (2011b). In the flood protecting problem, however, most of the decisions are binary variables determining whether and when a given measure (dike heightening, etc.) is applied. The first applications of ARO to integer recourse problems were Bertsimas and Caramanis (2007, 2010), and Vayanos et al. (2011), where the idea was to simply divide the uncertainty set into a dense grid of points and allow a different decision for each of them. Bertsimas and Georghiou (2015, 2014), and Hanasusanto et al. (2014) proposed to use specific decision rules for the integer variables whose 'shape' is optimized. However, these methodologies do not scale well with the size of the problem, which makes their application to problems like ours impossible.

The approach used in our paper is to construct multi-period adjustable integer decisions by means of multi-stage splitting of the uncertainty set into subsets. This is the approach taken in Postek and den Hertog (2016) and Bertsimas and Dunning (2016) where, having determined the splits of the uncertainty set, a different constant decision is applied for each part of the uncertainty set. The essence of this approach lies in determining the conditions that a splitting needs to satisfy in order to improve on the decisions' adaptivity. Since, however, in the model we consider it is not possible to determine the structure of the splits using the methods proposed by Postek and den Hertog (2016), our splitting strategy of the uncertainty set is pre-determined.

The structure of this paper is as follows. Section 6.2 gives the generic mathematical formulation of our problem without parameter uncertainty and adjustability. Section 6.3 defines how multi-stage parameter uncertainty is modeled and the corresponding adaptive decisions are constructed. Section 6.4 presents the results of numerical experiments for the Rhine - Meuse Estuary - Drechtsteden area.

6.2 Deterministic model

We consider the problem of constructing an optimal flood protection strategy. The aim of the strategy is to create a schedule when to take various measures that minimizes the present value of the measures' costs such that at each time moment and at each dike segment, the flood protection standards are satisfied. Hence, cost effectiveness is applied.

We assume that the flood protection system consists of N_s dike segments. The flood protection standards can be formulated in terms of *relative dike height* requirements - the height of the dike compared to the water level. The relative dike height can be improved by one of N_h dike heightenings of size $h \in \mathcal{H}$, and N_m large-scale measures such as, for example, changing the discharge distribution of a river (directing it via other river segments in the delta). Whereas a dike heightening affects only the relative dike height at a single segment, the large-scale measures affect more than one segment and its impact may differ throughout the time horizon after its implementation. We denote by $a_{m,s,\tau,t}$ the impact of implementing large-scale measure m at time τ on the relative dike height at segment s at time $t \geq \tau$. The flood protection standard requires that at each dike segment s and each time $t \in \{0, 1, \ldots, T\}$ the relative dike height is higher than or equal to the requirement $n_{s,t}$.

We define decision variables $x_{t,s,h} \in \{0, 1\}$ indicating whether the h-th dike heightening has been implemented at dike segment s at time t, and $y_{t,m} \in \{0, 1\}$ indicating whether large-scale measure m is implemented at time t. We denote the cost of dike heightening h for segment s by $p_{s,h}$, the cost of measure m by p_m, and we assume an inter-period discount rate $0 \leq d \leq 1$.

In this setting, the objective is to minimize the total discounted costs and the optimization problem is:

$$
\begin{aligned}
\min_{x,y} \quad & \sum_{t=1}^{T} \frac{1}{(1+d)^t} \left\{ \sum_{s=1}^{N_s} \sum_{h=1}^{N_h} p_{s,h} x_{t,s,h} + \sum_{m=1}^{N_m} p_m y_{t,m} \right\} & & \\
\text{s.t.} \quad & \sum_{\tau=0}^{t} \left\{ \sum_{h=1}^{N_h} h x_{\tau,s,h} + \sum_{m=1}^{N_m} a_{m,s,\tau,t} y_{t,m} \right\} \geq n_{s,t}, & \forall s, t & \quad (6.1a) \\
& \sum_{h=1}^{N_h} x_{t,s,h} \leq 1, & \forall t, s & \quad (6.1b) \\
& L_k(\boldsymbol{x}, \boldsymbol{y}) \leq 0, & \forall k = 1, \ldots, K & \quad (6.1c) \\
& x_{t,s,h}, y_{t,m} \in \{0, 1\}, & \forall t, s, h, m, &
\end{aligned}
$$

where the constraints are:

(6.1a) relative dike height constraint at segment s at time t - the total impact of the

dike heightenings and the large-scale measures up to time t accommodates for the requirement $n_{s,t}$,

(6.1b) for each dike segment s and time t, only one of N_h heightenings is implemented,

(6.1c) K linear constraints - these may involve, for example (i) conditions how many times can a given large-scale measure be taken, (ii) which measures cannot be taken together, (iii) redundant constraints which the solution must satisfy and which improve the solution time.

6.3 Modeling uncertainty and adjustability

6.3.1 Introduction

In the previous section we considered the relative dike height requirements $n_{s,t}$ to be a deterministic quantity known in advance. In fact, $n_{s,t}$ is based on the protection standards but also on the sea level rise:

$$n_{s,t} = n_{s,0} + \beta_s \sum_{\tau=1}^{t} r_\tau,$$

where $n_{s,0}$ is the initial requirement, r_τ is the sea level rise in time period $[\tau - 1, \tau]$, and β_s is the sensitivity of the dike segment s to the sea level rise - for example, sensitivity β_s of dike segments closer to the sea is higher than the dike segments situated higher along the rivers.

In this setting, the realized sea level rise may turn out to be different from the predicted values. Using only the forecast values of the sea level rise may lead to the solution becoming infeasible if small deviations from the given values occur (Ben-Tal et al. 2009). Secondly, even if the solution stays feasible, the future decisions are fixed and do not adjust to the outcomes of the uncertain sea level rise from the earlier periods. In the following, we discuss in detail the uncertainty structure w.r.t. the sea level rise and then develop the adjustable robust version of the deterministic model of Section 6.2.

6.3.2 Uncertainty structure

We divide the time horizon into intervals with points $0 = t_0 < t_1 < t_2 < \ldots < t_{J-1} < t_J = T$ and assume that in each of the intervals $[t_{j-1}, t_j]$, $j = 1, \ldots, J$ the sea level rise is equal to r_j which value is known only at time t_j. In this way (i) we allow the dynamics of the sea level rise to differ over time, (ii) it is possible to readjust the decisions at time points $t_1, \ldots, t_{J-1}$ after the value of the respective sea level rise is known.

Example 6.1 *Suppose that $T = 10$, and we define $J = 3$ with $t_1 = 3$ and $t_2 = 6$. We then have that the time horizon is divided into intervals:* $[0, 3]$, $[3, 6]$ *and* $[6, 10]$. *At time points t_1 and t_2 decisions can be readjusted, based on the revealed sea level rise speed in the periods* $[0, 3]$ *and* $[3, 6]$, *respectively.*

We assume that in each of the intervals $[t_{j-1}, t_j]$ the sea level rise belongs to an interval

$$r_j \in \mathcal{R}_j = [(1-\rho)\overline{r}_j, (1+\rho)\overline{r}_j], \quad j = 1, \ldots, J,$$

where $\overline{r}_j$ is the forecast value and parameter $0 \leq \rho \leq 1$ specifies the degree of uncertainty about r_j. For simplicity, we assume that $\overline{r}_j = \overline{r}$ for all j.

Let us denote the vector of uncertain sea level rises by $\boldsymbol{r} = (r_1, r_2, \ldots, r_J)$ and assume that the sea level rise speed (the sea level grows linearly throughout the interval) is constant in each of the intervals $[t_{j-1}, t_j]$. Then, we model the relative dike height requirement as:

$$n_{s,t}(\boldsymbol{r}) = n_{s,0} + \beta_s \left(\sum_{j: t_j \leq t} r_j + \frac{t - t_{j(t)}}{t_{j(t)+1} - t_{j(t)}} r_{j(t)} \right), \tag{6.2}$$

where $n_{s,0}$ is the initial requirement in period 1, r_j the rise of the sea level throughout the interval $[t_{j-1}, t_j]$, and $j(t) = \min\{j : t_j \geq t\}$, that is, $j(t)$ is the smallest j such that t belongs to the interval $[t_j, t_{j+1}]$. Formula (6.2) means that the requirement $n_{s,t}(\boldsymbol{r})$ is equal to the sum of the initial requirement, the sea level rises from all the full intervals that passed until time t and the sea level rise in the interval containing t up to this time moment, multiplied by the sensitivity β_s.

6.3.3 Adjustability of decisions - applicability of adaptive splitting

Adjustability of integer decisions consists in its essence in splitting the uncertainty set into subsets and designing different later-stage decisions for each of these subsets. For example, one splits the set $\mathcal{R}_1$ into two sets $\mathcal{R}_1^1 = [(1-\rho)\overline{r}_j, z]$ and $\mathcal{R}_1^2 = [z, (1+\rho)\overline{r}_j]$, such that $\mathcal{R}_1^1 \cup \mathcal{R}_1^2 = \mathcal{R}_1$ Then, at time t_1 when r_j is observed and it is known whether it holds that $r_1 \in \mathcal{R}_1^1$ or $r_1 \in \mathcal{R}_1^2$. Then, one can differentiate the decisions implemented from time t_j on, whereas there can be only one possible decision implemented before t_j, since it is not known yet what r_1 is.

One of the crucial questions in this setting is how to choose the point z where the sets $\mathcal{R}_j$ are split. Postek and den Hertog (2016) provide theoretical results on how to do it. This approach relies on the observation that for different constraints in the problem, different scenarios r_j are the worst-case scenarios, i.e., maximizing the left hand side. They prove that in order to improve on the worst-case objective function, these scenarios need to end up in different subsets $\mathcal{R}_1^1$ and $\mathcal{R}_1^2$ as a result of the

splitting procedure. The splitting procedure is then applied iteratively, resulting in a more and more adjustable decision structure.

Remark 6.1 *Here, we make an important distinction that by the worst-case scenario we mean a realization of the uncertain parameter within the specified uncertainty set that is least favorable to the given set of decisions. This should be opposed to the 'worst scenarios' considered in the water management community, corresponding to a specific sea level rise trajectory prescribed by, for example, the government.*

In the specific problem we consider, this method cannot be applied as (i) all the constraints of the problem depend on the uncertain parameter $\boldsymbol{r}$ in the same way, that is, the highest the sea level rise, (ii) the uncertainty set is a box, i.e. we have that $\mathcal{R} = \mathcal{R}_1 \times \ldots \times \mathcal{R}_J$. Combination of these two factors leads to the conclusions that it is always a single scenario $\boldsymbol{r}$ that maximizes the right-hand side term $n_{s,t}(\boldsymbol{r})$ in the safety requirement constraint. This claim is proved in the later part of this paper as Proposition 6.1 and it implies that we need (i) another methodology of choosing how to split the uncertainty set $\mathcal{R}$ than the one of Postek and den Hertog (2016), (ii) a different objective than simply the worst-case cost over the entire uncertainty set (total cost of the objective assuming that the flood protection policy is feasible for all $\boldsymbol{r} \in \mathcal{R}$. In the following subsections we outline the adjustability structure of our choice and demonstrate mathematically the need for a different objective function, which shall be the so-called *average worst-case objective*.

6.3.4 Adjustability of decisions - our approach

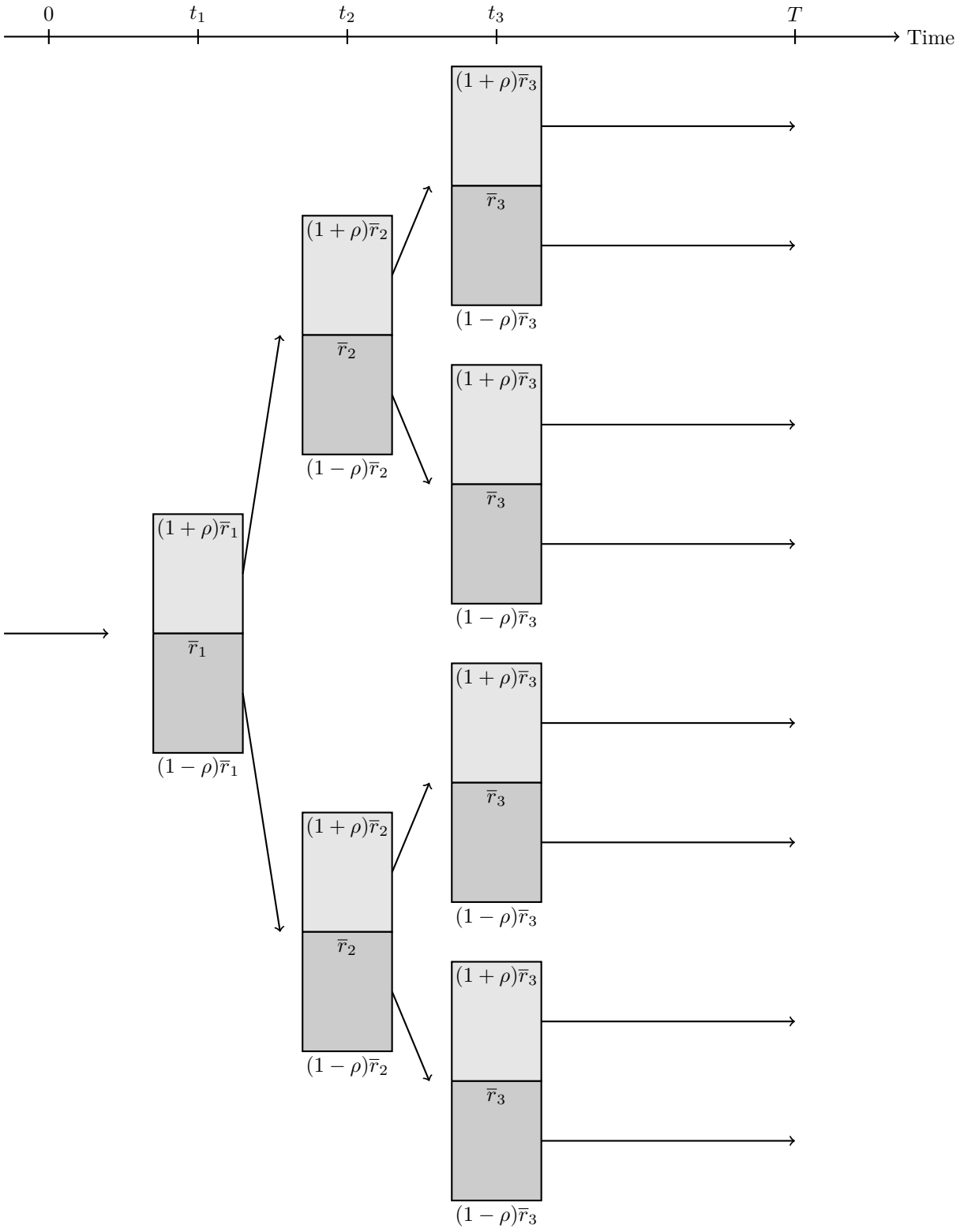
As for the splitting methodology, we assume that at $t_1, t_2, \ldots, t_j$, depending whether the realized value r_j is such that $r_j \in \mathcal{R}_j^1 = [(1-\rho)\overline{r}_j, \overline{r}_j]$ or $r_j \in \mathcal{R}_j^2 = [\overline{r}_j, (1+\rho)\overline{r}_j]$, different decisions can be implemented in the periods to come. We choose to split the uncertainty sets $\mathcal{R}_j^2$ at points $\overline{r}_j$ for its simplicity and lack of theoretical reasons justifying another splitting structure.

In other words, before time t_1 there is only one decision scenario to be implemented. From time t_1 and before t_2 there are two decision possibilities, between t_2 and t_3 four, and so on. This gives rise to the tree decision structure given in Figure 6.1.

As shown in Figure 6.1, $J-1$ split points result in a decision tree with 2^{J-1} *paths* - sequences of arrows from time 0 to T. We enumerate the paths using $(J-1)$-tuples $\boldsymbol{\alpha} \in \{1,2\}^{J-1}$ which is illustrated by the following example.

Example 6.2 *In Figure 6.1,* $\boldsymbol{\alpha} = (2,2,1)$ *corresponds to the decision paths that branches 'up' at times* t_1 *and* t_2*, and then 'down' at time* t_3*.*

Figure 6.1 – Adjustability scheme when $J = 4$. At times points t_1, t_2, t_3 it turns out whether the sea level rise belonged to $[(1-\rho)\overline{r}_j, \overline{r}_j]$ (symbolized by the dark gray part of the box) or $[\overline{r}_j, (1+\rho)\overline{r}_j]$ (symbolized by the light gray part of the box).



We define that for each of the decision paths for the entire planning horizon $[0, T]$ there is a separate decision vector $x_{t,s,h,p}(\boldsymbol{\alpha})$ and $y_{t,m,p}(\boldsymbol{\alpha})$, where $t = 0, \ldots, T$. Each of the decision paths is then uniquely identified with a trajectory of the sea level rise $\boldsymbol{r}$:

$$\alpha_j = \begin{cases} 1 & \text{if } r_j \in \mathcal{R}_j^1 \\ 2 & \text{if } r_j \in \mathcal{R}_j^2 \end{cases}, \quad j = 1, \ldots, J-1.$$

In a similar way, we construct multi-period uncertainty sets for $\boldsymbol{r}$ uniquely identifiable with $\boldsymbol{\alpha}$'s as:

$$\mathcal{R}(\boldsymbol{\alpha}) = \mathcal{R}_1^{\alpha_1} \times \mathcal{R}_2^{\alpha_2} \times \ldots \times \mathcal{R}_{J-1}^{\alpha_{J-1}} \times \mathcal{R}_{J-1}^{\alpha_{J-1}} \times \mathcal{R}_J.$$

The relation between the decisions $x_{t,s,h,p}(\boldsymbol{\alpha})$ and $y_{t,m,p}(\boldsymbol{\alpha})$ and the uncertainty sets $\mathcal{R}(\boldsymbol{\alpha})$ is that we require each of the decision paths to satisfy the relative dike requirement for each $\boldsymbol{r}$ belonging to an uncertainty set corresponding to that path (adjustable robust version of constraint (6.1a)):

$$\sum_{\tau=0}^{t} \left\{ \sum_{h=1}^{N_h} h x_{\tau,s,h}(\boldsymbol{\alpha}) + \sum_{m=1}^{N_m} a_{m,s,t,\tau} y_{t,m}(\boldsymbol{\alpha}) \right\} \geq n_{s,t}(\boldsymbol{r}), \quad \forall s, t \quad \forall \boldsymbol{r} \in \mathcal{R}(\boldsymbol{\alpha}). \quad (6.3)$$

However, as some of the decision paths from 0 to T coincide up to a certain time point, it is necessary to include *nonanticipativity constraints*, ensuring that decisions corresponding to paths that are the same up to some point, are also the same, to prevent that decisions are based on information not available yet. These constraints can be formulated as:

$$x_{t,s,h}(\boldsymbol{\alpha}') = x_{t,s,h}(\boldsymbol{\alpha}''), \quad y_{t,m}(\boldsymbol{\alpha}') = y_{t,m}(\boldsymbol{\alpha}''), \quad \forall t : t < t_j, \quad \boldsymbol{\alpha}'_{1:j} = \boldsymbol{\alpha}''_{1:j}, \quad (6.4)$$

where $\boldsymbol{\alpha}_{1:j}$ denotes a subvector of vector $\boldsymbol{\alpha}$ consisting of its first j components. Nonanticipativity constraints are illustrated by the following example.

Example 6.3 *In the scheme of Figure 6.1 we have that the paths corresponding to the two top branches of the tree correspond to* $\boldsymbol{\alpha}' = (2, 2, 2)$ *and* $\boldsymbol{\alpha}'' = (2, 2, 1)$*. Thus we have that* 2 *is the largest* j *such that it holds that* $\boldsymbol{\alpha}'_{1:j} = \boldsymbol{\alpha}''_{1:j}$*. This, combined with (6.4) implies that* $x_{t,s,h,p}(2, 2, 2) = x_{t,s,h,p}(2, 2, 1)$ *and* $y_{t,m,p}(2, 2, 2) = y_{t,m,p}(2, 2, 1)$ *for* $t < t_2$*.*

6.3.5 Objective function

In the deterministic problem (6.1) there is only a single decision path and no uncertainty so that the objective function is simple and corresponds to the total discounted costs of the measures taken.

In the adjustable case, however, each of the paths $\boldsymbol{\alpha}$ yields a different decision path that has a different total cost $C(\boldsymbol{\alpha})$:

$$C(\boldsymbol{\alpha}) = \sum_{t=0}^{T} \frac{1}{(1+d)^t} \left\{ \sum_{s=1}^{N_s} \sum_{h=1}^{N_h} p_{s,h} x_{t,s,h}(\boldsymbol{\alpha}) p_{s,h} + \sum_{m=1}^{N_m} p_m y_{t,m}(\boldsymbol{\alpha}) \right\},$$

and the objective function of the entire problem should somehow take into account the costs related to all the paths. A typical choice in Robust Optimization is to consider the worst-case cost as the objective function - in our case, the cost of the most expensive decision path. Suppose that indeed, for fixed $\boldsymbol{t} = (t_1, \ldots, t_{J-1})$ we minimize the worst-case objective, i.e. the value of the objective function under the assumption that the worst-possible realization $\boldsymbol{r} \in \mathcal{R}$ is realized. The optimization problem is then:

$$\min_{x,y} \max_{\boldsymbol{\alpha} \in \{1,2\}^{n_{\boldsymbol{z}}}} C(\boldsymbol{\alpha})$$

$$\text{s.t.} \sum_{\tau=0}^{t} \left\{ \sum_{h=1}^{N_h} h x_{\tau,s,h}(\boldsymbol{\alpha}) + \sum_{m=1}^{N_m} a_{m,s,\tau,t} y_{t,m}(\boldsymbol{\alpha}) \right\} \geq n_{s,t}(\boldsymbol{r}), \quad \forall s,t \quad \forall v \in \mathcal{R}(\boldsymbol{\alpha}) \tag{6.5a}$$

$$n_{s,t}(\boldsymbol{r}) = n_{s,0} + \beta_s \left(\sum_{j: t_j \leq t} r_j + \frac{t - t_{j(t)}}{t_{j(t)+1} - t_{j(t)}} r_{j(t)} \right) \tag{6.5b}$$

$$x_{t,s,h}(\boldsymbol{\alpha}') = x_{t,s,h}(\boldsymbol{\alpha}''), \quad y_{t,m}(\boldsymbol{\alpha}') = y_{t,m}(\boldsymbol{\alpha}''), \quad \forall t : t < t_j, \quad \boldsymbol{\alpha}'_{1:j} = \boldsymbol{\alpha}''_{1:j} \tag{6.5c}$$

$$\sum_{h=1}^{N_h} x_{t,s,h}(\boldsymbol{\alpha}) \leq 1, \quad \forall t, s, \quad \forall \boldsymbol{\alpha} \tag{6.5d}$$

$$L_k(\boldsymbol{x}(\boldsymbol{\alpha}), \boldsymbol{y}(\boldsymbol{\alpha})) \leq 0, \quad \forall k = 1, \ldots, K, \quad \forall \boldsymbol{\alpha} \tag{6.5e}$$

$$x_{t,s,h}(\boldsymbol{\alpha}), y_{t,m}(\boldsymbol{\alpha}) \in \{0,1\}, \quad \forall t, s, h, m, \quad \forall \boldsymbol{\alpha}. \tag{6.5f}$$

Let us denote by $\mathrm{WC}(\boldsymbol{t})$ the optimal value to (6.5) for a given $\boldsymbol{t}$. There is no adjustability if $J = 1$ and we denote the objective value in this case by $\overline{\mathrm{WC}}$.

As the following Proposition shows, it is not possible that $\mathrm{WC}(\boldsymbol{t})$ is better than $\overline{\mathrm{WC}}$.

Proposition 6.1 *For every $\boldsymbol{t}$ it holds that $WC(\boldsymbol{t}) = \overline{WC}$.*

Proof. It follows easily that $\mathrm{WC}(\boldsymbol{t}) \leq \overline{\mathrm{WC}}$ from the fact that every feasible solution to the problem with $J = 1$ is also feasible to the problem with $J > 1$. Assume there exists $\boldsymbol{t}$ such that $\mathrm{WC}(\boldsymbol{t}) < \overline{\mathrm{WC}}$. We have that the optimal solution corresponding to $\boldsymbol{\alpha}^* = (2, 2, \ldots, 2)$ is feasible for every

$$\mathcal{R}((2,2,\ldots,2)) = [\overline{r}_1, (1+\rho)\overline{r}_1] \times [\overline{r}_2, (1+\rho)\overline{r}_2] \times \ldots \times [\overline{r}_{J-1}, (1+\rho)\overline{r}_{J-1}] \times \mathcal{R}_J.$$

However, since the only constraint dependent on $\boldsymbol{r}$ is (6.5a), and its left hand side depends monotonically on components of $\boldsymbol{r}$, we have that the optimal solution corresponding to $\boldsymbol{\alpha}^*$ is feasible for all $\mathcal{R}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha} \in \{1,2\}^{J-1}$, and hence, for the entire uncertainty set:

$$\mathcal{R} = [(1-\rho)\overline{r}_1, (1+\rho)\overline{r}_1] \times [(1-\rho)\overline{r}_2, (1+\rho)\overline{r}_2] \times \ldots \times [(1-\rho)\overline{r}_{J-1}, (1+\rho)\overline{r}_{J-1}] \times \mathcal{R}_J.$$

In other words, decisions feasible for the uncertainty set being a product of the 'worse subintervals' for r_j (containing the higher values) is feasible also for the 'better intervals' and hence, for the entire uncertainty set. But this implies that the decision sequence corresponding to $\boldsymbol{\alpha}^*$ is also feasible for problem with $J = 1$ (no adjustability) and, at the same time, gives a better objective value than $\overline{\text{WC}}$. This is a contradiction with the assumption that $\overline{\text{WC}}$ is the best objective value for problem with $J = 1$. ■

Proposition 6.1 shows that no improvement in the objective value is possible if the objective is the worst-case cost. More generally, it can be proven in a similar fashion that no adjustability schemes can lead to an improvement of the worst-case formulation. A simple consequence of this fact is, as we have already signalized, that the methods of adaptive splits of Postek and den Hertog (2016), in which the division of the uncertainty set $\mathcal{R}$ is determined on the basis of solutions, cannot be applied. It is exactly for that reason that in our application the splits of the uncertainty sets $[(1-\rho)\overline{r}_j, (1+\rho)\overline{r}_j]$ into halves $[(1-\rho)\overline{r}_j, \overline{r}_j]$ and $[\overline{r}_j, (1+\rho)\overline{r}_j]$ is pre-determined.

As a consequence, we propose that instead of the worst-case cost of a decision path, the objective function is the average of the costs associated with all the paths:

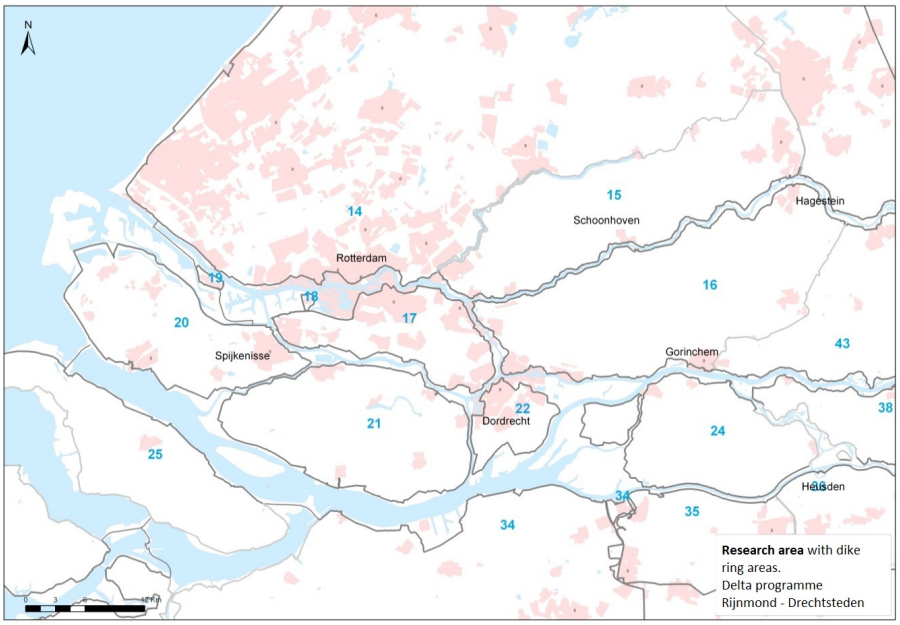
$$\frac{1}{2^{J-1}} \sum_{\boldsymbol{\alpha} \in \{1,2\}^{J-1}} C(\boldsymbol{\alpha}).$$

The weights $1/2^{J-1}$ correspond to the fact that each of the uncertainty subsets $\mathcal{R}(\boldsymbol{\alpha})$ has the same volume and that we assume both parts to be evenly likely (which happens if the underlying probability distribution is uniform). Secondly, due to the fact that each $C(\boldsymbol{\alpha})$ plays a role, it is almost sure that for the uncertainty subsets with smaller sea level increase, smaller later-stage investments are needed to meet the safety standards. In this way, we can expect increasing J to lead to better optimal values.

In the end, the adjustable problem we consider is:

$$\begin{aligned} \min_{x(\boldsymbol{\alpha}), y(\boldsymbol{\alpha})} \quad & \tfrac{1}{2^{J-1}} \sum_{\boldsymbol{\alpha} \in \{1,2\}^{J-1}} C(\boldsymbol{\alpha}) \\ \text{s.t.} \quad & \text{(6.5a) - (6.5f).} \end{aligned} \tag{6.6}$$

Figure 6.2 – The Rhine - Meuse Estuary - Drechtsteden area. Thick lines stand for the dike segments, numbers denote the dike rings - areas surrounded by dikes or high grounds from all sides.



6.4 Numerical experiment

6.4.1 Region

In our numerical experiment we consider the Rhine - Meuse Estuary - Drechtsteden area illustrated in Figure 6.2, consisting of $N_s = 150$ dike segments. For each dike segment $N_h = 22$ dike heightenings are possible of size $\{0.1, 0.2, \ldots, 2.0, 2.5, 3.0m\}$. Also there are $N_m = 14$ large scale water measures:

1. **Replacement of the Measlantbarrier.** This storm surge barrier is located at Hoek van Holland in the Nieuwe Waterweg. This barrier together with dikes and dunes protects the province Zuid-Holland against floods from sea. The storm surge barrier protects against high water and can be closed in emergencies. This measure can be implemented only from 2070 on.

2. **Improving the Maeslantbarrier** before 2070 and its replacement in 2070.

3. **Change the river discharge distribution via the IJssel river.** When the discharge via the Lek river becomes greater than $8000m^3/s$, more water is discharged via the IJssel river.

4. **Change the river discharge distribution via the Waal river.** When the discharge via the Lek river becomes between $8000m^3/s$ and $16000m^3/s$, more water is discharged via the Waal river.

5. **High water level Waal.** When the discharge of the Rhine is above $16000m^3/s$, more water will be discharged via the Waal river.

6. **Robert Koning.** In this measure extra space for the Merwede is created with a flood channel through the Land of Heusden and Altena. It is a green river, called Robert Koning, that is used when water levels exceed a certain height. It is expected that these high water levels occur once every 100 years.

7. **Room for the River packages (4 measures).** There are several measures that dig off areas around the river Merwede. The four measures that can be taken are the measures RvR klein 01, RvR klein 02*, MW007, and MW42_3. Two packages (RvR klein 01 and RvR klein 02*) combine several measures that dig off areas around Merwede.

8. **Grevelingen.** With this measure the possibility is created to store water at the Grevelingen lake.

9. **Spaargaren 1.** In this measure the seaside is closed by sluices at the Maeslantbarrier.

10. **Spaargaren 2.** Around the Nieuwe and Oude Maas rivers, sluices close the seaside with extra discharge capacity.

11. **Spaargaren 3.** The sluices close the seaside around the Nieuwe and Oude Maas rivers as Spaargaren 2, but without extra discharge capacity.

The big measures are illustrated in Figure 6.3.

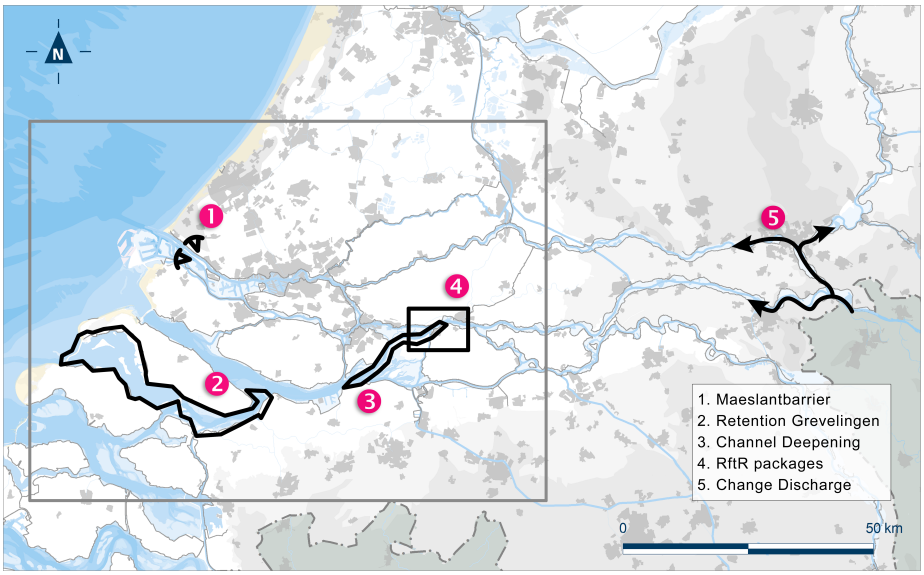
6.4.2 Piping problem

An additional trait of the problem we consider is the so-called *piping problem* affecting some of the dike segments. If groundwater flows wash away the sand below a dike, a pipe through the dike can arise (Deltares 2014). The flood probability increases enormously and the piping problem needs to be solved instantly in order to reduce the flood probability significantly. This issue cannot be addressed by lowering the water level, but only by strengthening the dikes. Let us denote by $\mathcal{S}^p$ the set of dike segments for which the piping problem exists.

The piping problem enters our problem in a way that for each $s \in \mathcal{S}^p$ it holds that (i) every dike heightening $h \in \mathcal{H}$ and (ii) solving the piping problem, have a fixed cost component c_s^f which is the same for both $p_{s,h}$ and r_s, such that:

$$p_{s,h} = c_s^f + p_{s,h}^v \qquad r_s = c_s^f + r_s^v,$$

Figure 6.3 – The Rhine - Meuse Estuary - Drechtsteden area with some of the large-scale measures. The box includes measures considered in the planning kit of Kind et al. (2014).



where $p^v_{s,h}$ is the variable cost of dike heightening depending on $h \in \mathcal{H}$, r_s is the total cost of solving the piping problem, and r^v_s is the remaining cost of solving the piping problem for dike segment s.

In our optimization model, we incorporate the piping issue in such a way that we assume that the piping problem, for dike segments s for which it exists, is solved at time $t = 0$. Thus, the fixed cost r_s is incurred at $t = 0$ for each $s \in \mathcal{S}^p$. In order to avoid double-counting of the fixed cost at $t = 0$ we assume that the cost of dike heightening h (if implemented) is equal to $p^v_{s,h}$ for that time period.

Since the total cost of solving the piping problem are relatively large compared to the cost of dike heightenings, in our numerical results we shall report on the total cost both with and without the cost of solving the piping problem.

6.4.3 Research questions and numerical results

The particular research questions that we subsequently answer in the coming sections, are as follows:

1. What is the nominal (nonrobust) solution and what the dangers are if uncertainty w.r.t. sea level rise is present?

2. What are the (adjustable) robust solutions, how do they change as more adjustability is allowed in the model?

3. How do the robust solutions differ from the nominal solution?

4. How do the robust solutions change when either the discount rate d or the the uncertainty level ρ changes?

5. How do the robust solutions change when an adjustability scheme different to that of Section 6.3 is applied? This question is considered, for reasons of brevity, in Appendix 6.A.

In our numerical experiment, the planning horizon is 2020 - 2200, with the possibility to implement decisions every 10 years, which gives 19 decision stages in total. When talking about the results, we shall often refer by $t = 0$ to year 2020, $t = 1$ to year 2030 and so on.

In our numerical experiment, we shall consider equidistant splitting points (if any adjustability is present) t_j, $j = 1, \ldots, J-1$, to be defined as

$$t_j = jW,$$

where W is the length of the interval between t_{j-1} and t_j as measured in the 'big' time intervals of 10 years, chosen to be independent from j. In this way, $W = 3$ corresponds to differentiating the decisions every $3 \cdot 10 = 30$ years. We solve the optimization model for different values of $J-1$ and W.

The sensitivities β_s and initial dike shortages $n_{s,0}$ are estimated from the data. For the robust models we assume that there is $\rho = 40\%$ uncertainty about the sea level rise speed, motivated by the Dutch KNMI uncertainty intervals for sea level rise speed.[1] We also solve a nonrobust (nominal) model where $\rho = 0\%$. The discount rate is assumed to be equal to $d = 5.5\%$. We assume that a dike segment should not be heightened more often than once in 30 years, which we denote as the *frequency constraint*.

Due to the size of the model and the number of binary variables, it is not expected that it is solved to full optimality within reasonable time limits. However, a particular feature of the model is that once the decisions w.r.t. the large-scale measures are fixed, the optimization problem related to the dike heightenings per each dike segment can be solved separately. For that reason, each time we reach the sub-optimal solution to the entire optimization model, we run an extra step consisting in re-optimization of the dike decisions, calling it the *dike reoptimization step*.

The optimization model formulation has been coded using the C++ language and is solved with the Gurobi solver (Gurobi Optimization 2015). For each instance with given W and J we set time limit of 10h and the MIP optimality gap to 0.1%. Problems have been implemented using 16-core computing nodes of the Lisa computing cluster at SURF SARA computing center in Amsterdam.

[1] See the document (in Dutch) 'Zeespiegelveranderingen in de toekomst', at http://www.knmi.nl/kennis-en-datacentrum/achtergrond/zeespiegelveranderingen-in-de-toekomst.

Table 6.1 – Results for the nominal solution. 'Obj value' - the optimal value (in millions of euros) of the optimization problem without piping cost; 'Opt. gap' - the MIP optimality gap after 10 hours of computation with dike reoptimization (without dike reoptimization); 'Dike cost $t = 0$' - total cost of dike heightenings implemented at $t = 0$

Nominal			
Obj value	Opt. gap	Obj value + piping	Dikes cost $t = 0$
1004.71	0.1% (0.1%)	3607.50	587.93

Table 6.2 – Implementation schedules of the used large-scale measures for the nominal solution.

Measure	Implementation year
Replacing the Maeslantbarrier	2180
Grev100	2090
MW42 3	2060
MW007	2020

6.4.3.1 Nominal solution and its shortcomings

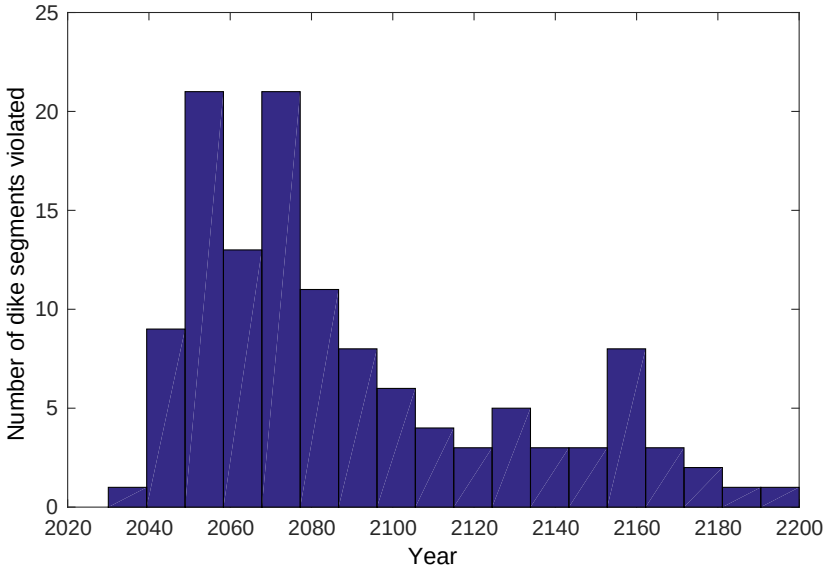
Table 6.1 presents the results for the nominal solution, i.e. assuming that there is no uncertainty about the sea level rise, and Table 6.2 presents the corresponding implementation schedule of the large-scale measures. We observe that the model has been solved nearly to optimality, attaining the total cost of 1004 (excluding the piping cost). Of all the variable cost, the largest share are the here-and-now dike heightening cost, 587. We can clearly see that the fixed cost of solving the piping problem (2602) dominate the cost of the entire solution, as they raise the total cost to the level of 3607.

Regarding the large-scale measures, in Table 6.2 we observe that only four of them are used: 'Replacing the Maeslantbarrier', 'Grev100', 'MW42 3' and 'MW007'. This gives an indication that not all of the measures are relevant and are worth maintaining their readiness to use. This indication shall be confirmed in later experiments.

To investigate 'how bad the nominal solution can become', we check first for how many dikes and when are the safety standards violated if the sea level rise turns out to be the worst +40% value taken into account for the robust solutions. It turns out that throughout the planning horizon, the safety requirements are violated for 123 out of 150 dike segments. Figure 6.4 includes the histogram of the first time moments that for these 123 segments the safety standards are violated. We see that there is a 'peak' of the frequency of dike violations in time frame 2040-2080. On average, the first violation occurs in year 2087.

As an additional illustration of the advantages of taking into account the robustness and adjustability, we provide one more experiment. In this experiment, we shall begin with the nominal solution and re-optimize its decisions in the subsequent time

Figure 6.4 – Histogram of the first time moments of safety standards' violation for the 123 dike segments for which there is a violation (out of the total number of 150 dike segments), with +40% sea level rise increase compared to the nominal scenario.



periods to match the worst +40% value, so that it can be considered as an 'on-the-spot fixing' of the nominal solution. To be precise, the procedure is as follows.

1. At time $t = 0$ an optimal strategy is constructed with the assumption of no uncertainty in the sea level rise (using thus $\rho = 0$) and the here-and-now decisions are implemented.

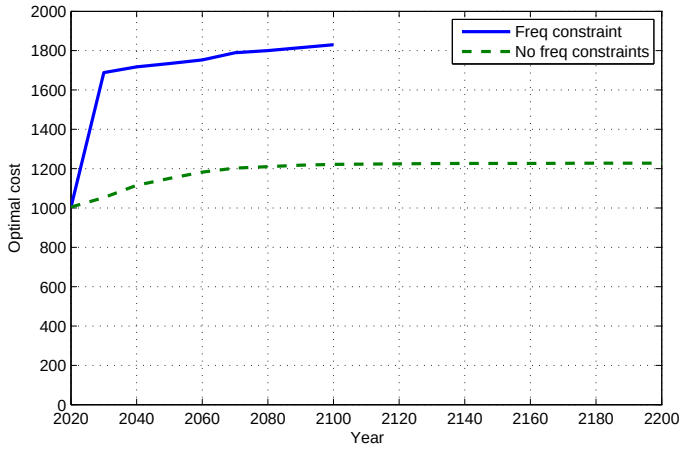
2. However, at $t = 1$ the sea level rise between t_0 and t_1 turns out to be the highest possible one with +40% and at time t_1 a new strategy is constructed (using again $\rho = 0$) for time horizon $[t_1, T]$ and the here-and-now decisions for t_1 are implemented.

3. Step 2 is repeated in time periods $[t_1, t_2]$, $[t_2, t_3]$,..., $[t_{J-1}, t_J]$.

It is important to note that in this experiment it is possible (and it will occur) that for certain dikes segments the safety standards are violated. Another issue is that the constraint that a dike segment is heightened at most once in 30 years might not be possible to be satisfied.

Indeed, we observe that the problem to solve becomes infeasible in period $t = 9$ (year 2110), due to the necessity to increase the height of certain dikes immediately while not allowed because of the frequency constraint. Strikingly, already the solution of the optimization problem solved at $t = 1$ (year 2030, the 'on-the-spot fixing' model is still feasible) involves a major increase in the total cost - a value of 1686, 68% up

Figure 6.5 – Increase of the total cost when the optimal solution is re-optimized due to the highest possible sea level rise realization. The continuous line corresponds to the case with frequency constraint (model infeasible in 2110) and without these constraints after 2020.



compared to the value 'promised' by the nominal solution in Table 6.1. The evolution of the optimal cost from the optimization problem solved at each stage is given by the continuous line in Figure 6.5.

Because it is not possible to 'keep fixing the nominal solution' without heightening the dikes more often than once in 30 years, we implemented also a variant of this test without the frequency constraint at stages later than $t = 0$. It turns out that the total cost is equal to 1227, that is, 22% higher (see Figure 6.5) than assumed in Table 6.1. Also, in the course of such a strategy, 7 of the 14 large-scale measures are used - an increase from 4 measures assumed in the initial solution.

From this test we see that in presence of uncertainty, even if the nominal solution is allowed to be 'fixed' later on, the cost of this fixing may be substantial. As we shall see in the section to come, the total cost of the 'fixed' solution is higher than the cost of the robust solution, i.e., the one that takes uncertainty into account at the very beginning.

6.4.3.2 Robust solutions

Having studied the disadvantages of the nominal solution, we turn to the (adjustable) robust solutions with sea level rise uncertainty $\rho = 40\%$. We begin the discussion of the results with the changes in the large-scale measures implementations for different solutions. We show the solutions for the nominal model and the robust models for $W = 2$ with zero and two splits in Table 6.3. We see from it (second row) that for the non-adjustable solution the measure 'RvR kl01' is not used at all, whereas for the adjustable solution its year of implementation varies between 2090 and 2130. The most stable measures are 'MW42 3' implemented by all solutions in 2050, and

Table 6.3 – Implementation schedules of the used large-scale measures for the nominal solutions and for the robust solutions with $W = 2$ and $J - 1 \in \{0, 2\}$. There is only one scenario for the nominal solution and for the robust solution with $J - 1 = 0$ (no adjustability) and four scenarios for $J - 1 = 2$.

	Nominal	$J - 1 = 0$ (no adjustability)	$W = 2$ and $J - 1 = 2$			
Scenario	-	-	1	2	3	4
Replacing the Maeslantbarrier	2180	2150	2110	2110	2110	2100
RvR kl01	-	-	2110	2130	2100	2090
Grev100	2090	2070	2080	2080	2080	2070
MW42 3	2060	2050	2050	2060	2050	2050
MW007	2020	2020	2020	2020	2020	2020

'MW007' which is implemented right away at the beginning of the planning horizon at 2020.

The cost and dike heightening implementation characteristics of the robust solutions are given in Table 6.4 (including also the nominal solution for ease of comparison) and we comment on it in a column-by-column basis, beginning the discussion on the example of models with $W = 2$, comparing it later to the nominal solution. Note that for all values of W the first row (solution with $J - 1 = 0$) is the same - this is because this is the solution with no adjustability.

In the first column, we can see that, as the degree of adjustability grows with the number of splits, the objective function value decreases. Already for one split we observe a drop of -3.73%, equivalent to approximately 45 million euro. For most adjustable solutions in Table 6.4 with three splits we have a decrease in the average worst-case cost at the level of -6.35%. The second columns provides the information about the remaining optimality gap of the MIP solver. For example, for three splits we can observe a remaining optimality gap 0.43% which without the reoptimization step would have been equal to 0.52%. Roughly speaking, this means that there is uncertainty whether the given solution can be improved by another 4.4 million euro.

In the third column the objective value of the optimization problem is added to the fixed cost of solving the piping problem, which dominate the cost structure. We can see that then, the relative improvements in the total cost due to adjustability oscillate between 1–2%.

The fourth and fifth columns provide information about the worst-case cost for the best and the worst $\boldsymbol{\alpha}$. That is, they provide the range to which the worst-case cost fall depending on the trajectory of the sea level rise. For the no adjustability case there is no gap between the two as there is only one scenario. But already for the single split solution, we can see that the 'best scenario' gives worst-case cost of 1106 compared to 1206 of the worst scenario. The difference between the two, equal to 100 million euro, informs us about the uncertainty in the total cost due to uncertainty

about the sea level growth. This difference takes about $100/1200 \approx 8.3\%$ of the total variable cost for $J - 1 = 1$. For the three splits solution we can see that the difference between the worst and best scenarios is even bigger, with values 1044 and 1207, respectively.

An important observation is that the worst-case scenario cost does not change that much for different number of splits, showing that minimization of the average worst-case objective does not lead to the minimization of the optimistic scenarios at the expense of much bigger riskiness, so that their average would go down.

The last three columns inform on the changes in the here-and-now dike heightening decisions. We can see that with adjustability, the decisions are different for 11-12 dike segments, with the average difference oscillating between 0.1–0.2m. The resulting change in the here-and-now dike heightening cost is equally small, compare for example (last column) 606 for three splits and 606 for no split.

With respect to the impact of different values of W we can observe that $W = 2$ and $W = 3$ provide slightly higher improvements in the total cost than $W = 4$. This can be linked to the fact that if the first differentiation of the total costs is done only after 40 years, then the only later-stage decisions that different for scenarios $\boldsymbol{\alpha}$ are the ones with discount factor at least $1.055^{-40} \approx 0.1175$. That is, the differentiation of decisions comes 'too late' to lead to substantial changes in the total cost.

6.4.3.3 Comparison of the robust and nominal solutions

As for the nominal solution, we can clearly see in Table 6.4 that the objective function value in the first column is significantly smaller than for the robust models: compare 1004 for the nominal solution to 1201 for the robust solutions without adjustability. We note however, that if the nominal solution is to be fixed 'on the spot' later, the cost grows sharply beyond the cost of the worst scenarios for the robust solution, as illustrated in Figure 6.5.

Due to lack of the need to accommodate for the uncertain sea level rise, the here-and-now investment decisions for the dikes differ for 20 dike segments, with the average difference of 0.19m (lower heightenings), with a total cost of 587 at $t = 0$, instead of 606 for the robust solution - about 3% less than the robust solution. In Table 6.3 we see that the nominal solution applies all the large-scale measures at least as late as the robust solutions.

Overall, we can say that incorporating 40% uncertainty in the sea level rise leads only to an about 4% increase of the expected total worst-case cost (when the fixed piping costs are included, about 12% when only variable costs are considered), and a 3-4% change in the here-and-now cost. These numbers seem to be rather small compared to the guarantees gained - recall what happens if the nominal solution is

Table 6.4 – Results for different values of W. 'Obj value' - the optimal value (in millions of euros) of the optimization problem without piping cost (change w.r.t. no adjustability in %); 'Opt. gap)' - the MIP optimality gap after 10 hours of computation with dike reoptimization (without dike reoptimization); 'Best scenario' - worst-case cost for the best scenario $\boldsymbol{\alpha}$; 'Worst scenario' - worst-case cost for the worst scenario $\boldsymbol{\alpha}$; '# dikes different' - number of dikes for which the here-and-now decision is different from the non-adjustable robust solution; 'Dike difference $t = 0$' the average differences of dike heightenings at $t = 0$ compared to the non-adjustable robust solution (for dikes for which the decision differs); 'Dike cost $t = 0$' - total cost of dike heightenings implemented at $t = 0$

Nominal								
Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dikes cost $t=0$
-	1004.71	0.1% (0.1%)	3607.50	-	-	20	0.19	587.93
Robustness with $W = 2$								
Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dikes cost $t=0$
0 (no adjustability)	1201.36	0.11% (0.11%)	3804.36	1201.35	1201.35	-	-	606.88
1	1156.57 (-3.73%)	0.20% (0.23%)	3759.57 (-1.18%)	1106.98	1206.16	11	0.14	606.59
2	1133.33 (-5.66%)	0.29% (0.32%)	3736.33 (-1.79%)	1061.78	1208.06	11	0.15	604.46
3	1125.08 (-6.35%)	0.43% (0.52%)	3728.08 (-2.01%)	1044.49	1207.61	11	0.15	606.39
4	1121.92 (-6.61%)	0.88% (0.97%)	3724.92 (-2.09%)	1037.25	1207.91	10	0.14	606.82
Robustness with $W = 3$								
Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dikes cost $t=0$
0 (no adjustability)	1201.35	0.1% (0.1%)	3804.35	1201.35	1201.35	-	-	606.87
1	1147.68 (-4.46%)	0.1% (0.1%)	3750.68 (-1.43%)	1092.16	1203.20	9	0.16	603.42
2	1135.06 (-5.84%)	0.22% (0.23%)	3738.06 (-1.77%)	1062.46	1205.60	10	0.15	602.91
3	1132.56 (-6.07%)	0.35% (0.37%)	3735.56 (-1.84%)	1058.00	1206.23	9	0.16	603.05
4	1131.97 (-6.12%)	0.43% (0.46%)	3734.97 (-1.85%)	1056.32	1206.95	9	0.16	602.79
Robustness with $W = 4$								
Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dikes cost $t=0$
0 (no adjustability)	1201.35 (0.00%)	0.1% (0.1%)	3804.35	1201.35	1201.35	-	-	606.87
1	1154.88 (-3.87%)	0.1% (0.1%)	3757.88 (-1.22%)	1105.56	1204.12	5	0.55	602.42
2	1147.96 (-4.44%)	0.13% (0.14%)	3750.96 (-1.40%)	1092.40	1204.43	5	0.65	602.17
3	1147.36 (-4.49%)	0.20% (0.20%)	3750.36 (-1.42%)	1091.05	1204.78	5	0.65	602.17
4	1147.23 (-4.51%)	0.19% (0.19%)	3750.23 (-1.42%)	1091.09	1204.84	5	0.65	602.17

to be fixed later.

6.4.3.4 Robust solutions when the key parameters change

Conclusions of the experiments up to now rely on the value of the two crucial parameters: (i) the discount rate d, and (ii) the uncertainty parameter ρ, whose values can be seen as somewhat arbitrary. We want to validate our results by considering the model with changes in these parameters and investigate what happens to the results of Table 6.4 if we change the discount rate from 5.5% to 2.5%, or change the uncertainty about the sea level rise from 40% to 60%. To make this comparison concise, we focus only on the models with $W = 3$. Table 6.6 presents the results.

In the first part of the table we can see the solutions for the changed discount rate. Lower discount rate implies that (i) the overall discounted cost grows higher, (ii) the cost of later-stage decisions grow relatively to the here-and-now decisions and hence, adjustability should become more important. This is indeed visible, for example, in the number of the dike segments for which the adjustable solutions involve different here-and-now decisions, which is 14 for a single split and 22 for two splits. This contrasts with the case $d = 5.5\%$ for which only 11–12 dike segments required different decisions because of the adjustability. Surprisingly, the adjustable decisions in this case require smaller here-and-now investment cost, which is reflected in the last column.

Also, we would like to highlight here the impact of the discount rate itself - the nonadjustable solutions of the first part of Table 6.6 implies a different here-and-now decision for 43 dike segments compared to the nonadjustable solution of Table 6.4, with an average difference of 0.42m, typically implying a larger here-and-now heightening. This is logical as with the lower discount rate, the here-and-now decisions become relatively cheaper.

The second part of the table includes the solutions for $\rho = 60\%$. Here, only four dike segments require a different here-and-now decision when adjustability is applied. Also, in the last column we see that in this setting the cost of the here-and-now dike heightening decisions are lower - 620 of $J - 1 = 1$ compared to 632 of $J - 1 = 0$.

6.4.3.5 Numerical experiment - results overview

Here, we summarize the conclusions of the numerical experiment of Section 6.4 including direct links to the tables and figures. These are as follows:

- not taking the uncertainty into account and assuming a ‘moderate’ sea level rise speed leads to solutions for which the safety standards are mostly violated within the period of about 40 years (see Figure 6.4) if the worst-case happens;
- not taking the uncertainty into account and assuming a ‘moderate’ sea level,

Table 6.5 – Results for changed discount rate (top part), uncertainty level (middle part), and (splitting strategy). Terminology the same as in Table 6.4.

$W = 3$ and discount rate 2.5%

Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dikes cost $t=0$
0 (no adjustability)	2291.73	0.92% (0.92%)	4894.73	2291.73	2291.73	0	0.0	940.46
1	2203.23 (-3.86%)	2.56% (2.68%)	4806.23 (-1.80%)	2109.58	2296.88	14	0.19	919.43
2	2146.97 (-6.31%)	3.52% (4.29%)	4749.97 (-2.95%)	1990.36	2305.81	22	0.2	902.84

$W = 3$ and uncertainty 60%

Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dikes cost $t=0$
0 (no adjustability)	1298.21	0.09% (0.10%)	3901.21	1298.21	1298.21	-	-	632.21
1	1226.74 (-5.50%)	0.10% (0.10%)	3829.74 (-1.83%)	1151.84	1301.59	4	0.2	620.98
2	1207.61 (-6.97%)	0.19% (0.20%)	3810.61 (-2.32%)	1111.67	1302.09	4	0.2	620.98
3	1204.05 (-7.25%)	0.29% (0.30%)	3807.05 (-2.41%)	1104.14	1302.89	4	0.2	620.98
4	1203.16 (-7.32%)	0.39% (0.40%)	3806.16 (-2.43%)	1102.50	1302.55	4	0.8	620.98

combined with 'on the spot' fixing of the situation leads to rapidly growing total cost of the solution and/or violating some of the constraints (safety requirements and *frequency constraints*, stating how often can a single dike segment be heightened), see Figure 6.5. In fact, the total cost of the 'on the spot problem fixing' strategy are higher than the maximum cost of the strategy taking into account robustness from the very beginning;

- considering 40% uncertainty in the sea level rise results in less than 10% difference in the expected total project cost (see Table 6.4), equivalent to about 130 million euro;
- adjustability allows for lower-cost solutions particularly when the realized uncertain sea level rise in the early periods is low, see Tables 6.4 and 6.6;
- adjustability and here-and-now decisions for dike segments play a much more important role when the discount rate is low (see the first part of Table 6.6, where 22 dikes require a different here-and-now decision, compared to 10 dike segments in Table 6.4);
- only a small part of the large-scale measures (5 out of 14) are used by the optimal solutions and their moment of implementation is relatively stable (see Table 6.3);
- a different set splitting strategy for adjustability (see Appendix 6.A), based on sums of sea level increase, does not change the solutions significantly.

6.5 Conclusion

In this paper we have considered the problem of finding the optimal flood protection strategy in a long-term horizon. In particular, the aim was to address two challenges: (i) one related to explicitly taking into account of the uncertainty related to the future sea level rise and (ii) the second one, requiring the future decisions to adjust to the revealed uncertainties from earlier periods. Both challenges have been addressed by applying the tools of integer-adjustable robust optimization as proposed in Postek and den Hertog (2016).

The numerical experiments reveal that taking uncertainty into account early leads to solutions that, compared to the nominal solutions based on 'moderate' sea level rise scenario, do not require expensive adjustments at later time stages, if the uncertainty deviates from the moderate scenario. At the same time, adjustability of decisions ensures that cheaper decisions are made when the realized uncertainty is low.

Also, taking into account uncertainty results in different here-and-now decisions for some of the dike segments, which facilitate better adjustments in the future. This

becomes even more important when the assumed discount rate is low (and the relative costs of the later-stage decisions are higher). In the end, we conclude that only a small part of the available large-scale measures are used in the optimal solutions.

Regarding future research, it is important to consider the problem of finding an optimal solution on a larger-scale region including also dike segments whose safety standards are driven by the (uncertain) river flow rather than by the sea. Furthermore, the model can be extended by taking into account costs arising in case a certain area is flooded.

Appendix

6.A Experiment - alternative splitting proposal

The results for the adjustable robust solutions up to now have been based on the splitting methodology presented in Section 6.3. In it, we have been differentiating the decisions based on whether r_1 fell into the upper or lower half of the corresponding interval, and the same for the later r_j.

Alternatively, one can argue that the splitting (adjustability) at time t_j should be based on the outcome of the sum $\sum_{s=1}^{j} r_s$. This is because the total water level depends on the sum of the sea level rises in the subsequent time periods rather than on each of them separately. Each of the uncertainty subsets $\mathcal{R}(\boldsymbol{\alpha})$ is identified then, instead with the sequence of inequalities in the following sequence (as in Section 6.3)

$$r_1 \ (\geq \text{ or } \leq) \ \bar{r}_1, \quad r_1 \ (\geq \text{ or } \leq) \ \bar{r}_2, \quad r_3 \ (\geq \text{ or } \leq) \ \bar{r}_2, \quad \ldots$$

with the sequence of inequalities in the following sequence:

$$r_1 \ (\geq \text{ or } \leq) \ \bar{r}_1, \quad r_1 + r_2 \ (\geq \text{ or } \leq) \ \bar{r}_1 + \bar{r}_2, \quad r_1 + r_2 + r_3 \ (\geq \text{ or } \leq) \ \bar{r}_1 + \bar{r}_2 + \bar{r}_2, \quad \ldots$$

Again, the objective function is a weighted average of the cost corresponding to each of the 2^{J-1} scenarios. However, in this setting it is no longer true that all uncertainty subsets $\mathcal{R}(\boldsymbol{\alpha})$ are equally probable if one assumes uniform probability distribution of $\boldsymbol{r}$ over $\mathcal{R}$ instead, the volume of each of the uncertainty subsets needs to be computed. In other words, with $J - 1 = 2$ the probability of the scenario $\boldsymbol{\alpha} = (2, 2)$ is equal to the proportion $p(\boldsymbol{\alpha})$ of the volume of the polytope

$$\mathcal{R}(\boldsymbol{\alpha}) = \{\boldsymbol{r} : \ (1-\rho)\bar{r}_j \leq r_j \leq (1+\rho)\bar{r}_j, \ j = 1, 2, 3, \ r_1 \geq \bar{r}_1, \ r_1 + r_2 \geq \bar{r}_1 + \bar{r}_2, \}$$

to the volume of the entire uncertainty set $\mathcal{R}$. We have computed these proportions $p(\boldsymbol{\alpha})$ using a Monte Carlo simulation and we note here only that these proportions (probabilities) give relatively more weight to the extreme uncertainty subsets - the ones where the sums of sea level rises is consistently in the 'lower part' or 'upper part', respectively.

Table 6.6 – Results for the changed splitting strategy. Terminology the same as in Table 6.4.

$W = 3$ and uncertainty 60% and alternative splitting strategy

Splits ($J-1$)	Obj value	Opt. gap	Obj value + piping	Best scenario	Worst scenario	# dikes different $t=0$	Dike difference $t=0$ (m)	Dike costs $t=0$
0 (no adjustability)	1201.35	0.1% (0.1%)	3804.35	1201.35	1201.35	-	-	606.87
1	1147.68 (-4.46%)	0.1% (0.1%)	3750.68 (-1.43%)	1092.16	1203.20	9	0.16	603.42
2	1132.08 (-5.77%)	0.15% (0.16%)	3735.08 (-1.82%)	1062.20	1204.34	8	0.16	603.15
3	1128.94 (-6.03%)	0.26% (0.27%)	3731.94 (-1.90%)	1056.84	1203.55	6	0.18	604.08
4	1128.35 (-6.08%)	0.33% (0.34%)	3731.35 (-1.92%)	1055.84	1203.62	7	0.17	603.94

The results of the experiment with a changed splitting strategy are given in the last part of Table 6.6. Obviously, since the objective function is weighted with different probabilities than in the original case, we should rather compare the new splitting strategy based on the last five columns. Regarding the worst-case costs in the worst and best scenario, they are nearly the same as the ones for $W = 3$ in Table 6.4, with the worst-scenario only slightly better than in Table 6.4. This difference is most likely accountable to the new probabilities assigning more weight to the extreme scenarios. We also observe that the number of dikes with different $t = 0$ decisions is rather stable, between 6 and 9. Overall, we conclude that in this parameter setting a different splitting strategy does not lead to significantly different solution, especially in the first time period.

Bibliography

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

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

D. Bertsimas and C. Caramanis. Adaptability via sampling. In *46th IEEE Conference on Decision and Control*, pages 4717–4722. IEEE, 2007.

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

D. Bertsimas and I. Dunning. Multistage robust mixed integer optimization with adaptive partitions. *Operations Research*, 64(4):980–998, 2016.

D. Bertsimas and A. Georghiou. Binary decision dules for multistage adaptive mixed-integer optimization. http://www.optimization-online.org/DB_FILE/2014/08/4510.pdf, 2014.

D. Bertsimas and A. Georghiou. Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Operations Research*, 63(3):610–627, 2015.

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

D. Bertsimas, D. A. Iancu, and P. A. Parrilo. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12): 2809–2824, 2011b.

R. Brekelmans, D. den Hertog, K. Roos, and C. Eijgenraam. Safe dike heights at minimal costs: the nonhomogeneous case. *Operations Research*, 60(6):1342–1355, 2012.

X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2012.

X. Chen, M. Sim, and P. Sun. A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071, 2012.

Deltares. Present dike standards must focus more on piping (in Dutch). http://www.deltares.nl/en/news/news-item/item/14978/present-dike-standardsmust-focus-more-on-piping, 2014.

C. Eijgenraam, J. Kind, C. Bak, R. Brekelmans, D. den Hertog, D. Duits, K. Roos, P. Vermeer, and W. Kuijken. Economically efficient standards to protect the Netherlands against flooding. *Interfaces*, 44(1):7–21, 2014.

C. Eijgenraam, C. Roos, R. Brekelmans, and D. den Hertog. Optimal strategies for flood prevention. *Management Science*, 2016.

V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483, 2014.

J. Gauderis, J. Kind, and R. van Duinen. Robustness of economically efficient flood protection standards: Monte Carlo analysis on top of cost-benefit analysis. In F. Klijn and T. Schweckendiek, editors, *Comprehensive flood risk management*, pages 851–858. Taylor & Francis Group, London, London, 2013.

Gurobi Optimization. Gurobi optimizer reference manual. http://www.gurobi.com, 2015.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. K-adaptability in two-stage robust binary programming. *Operations Research*, 63(4):877–891, 2014.

A. Jeuken, J. Kind, and J. Gauderis. Cost-benefit analysis of flood protection strategies for the Rhine - Meuse delta. Comprehensive flood risk management. Research for policy and practice. In F. Klijn and T. Schweckendiek, editors, *Proceedings of the 2nd European Conference on Flood Risk Management*, page 436, London, 2013. CRC Press, Taylor & Francis Group.

J. Kind. Economically efficient flood protection standards for the Netherlands. *Journal of Flood Risk Management*, 7:103–117, 2014.

J. Kind, R. Vos, A. Tijssen, A. Jeuken, and N. Slootjes. Towards the development and evaluation of adaptive flood risk management strategies for the Rhine Estuary - Drechtsteden. http://www.abrh.org.br/icfm6/proceedings/papers/PAP014707.pdf, 2014.

K. Postek and D. den Hertog. Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing*, 28(3): 553–574, 2016.

J. Van Alphen. The Delta Programme and updated flood risk management policies in the Netherlands. *Journal of Flood Risk Management*, 2015. doi: 10.1111/jfr3.12183.

D. Van Dantzig. Economic decision problems for flood prevention. *Econometrica*, 24(3): 276–287, 1956.

H. Van der Most, I. Tánczos, K. M. de Bruijn, and D. Wagenaar. New, risk-based standards for flood protection in the Netherlands. http://www.deltares.nl/app/uploads/2014/12/ICFM6-PAP014759-Flood-protection-standards-def1.pdf, 2014.

Ph. Vayanos, D. Kuhn, and B. Rustem. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In *2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, pages 7368–7373, 2011.