Primal and dual approaches to 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

vrijdag 19 januari 2018 om 10.00 uur
door

Franciscus Johannus Cornelius Tarcisius de Ruiter

geboren op 18 december 1990 te Culemborg.

Promotiecommissie:

Promotores:	prof.dr.ir. Dick den Hertog
	prof.dr. Dimitris Bertsimas
Copromotor:	dr. Ruud Brekelmans
Overige leden:	prof.dr. Erick Delage
	prof.dr. Wolfram Wiesemann
	prof.dr. Monique Laurent
	prof.dr. Kuno Huisman

Primal and dual approaches to adjustable robust optimization

Acknowledgements

The work presented in this thesis could not have been successful without the love and support from many, from inside academia, to friends and family.

First and foremost, I would like to single out the person who enthused me for research and guided me throughout my PhD. Dick den Hertog, I cannot possibly imagine a better supervisor than you have been. Beside aiding me with my research, you selflessly allowed me to travel and visit other research institutes and independently collaborate with other people. You are one of the most ethical and sincere researchers in the field. Your approach acts as an example that I can remind me of for the rest of my life.

I am greatly indebted to my second supervisor Dimitris Bertsimas. Staying at MIT in my first year, as well as inviting me for a second stay, gave me an invaluable experience. Every meeting we discussed 'new and exciting' ideas. You stimulated me to have a positive outlook, pursue various research ideas and aim for the best. I am forever thankful for the opportunity you gave me to stay at MIT and the things I have learned.

I also express gratitude to my copromotor, Ruud Brekelmans for carefully reading parts of my thesis and his help with the work on the last chapters in the early days of my PhD research.

I would like to thank my committee members for carefully reading and providing valuable comments on an earlier version of my thesis, as well as the positive encounters we had during my PhD. With Wolfram Wiesemann I have attended my very first conference (APMOD, Warwick 2014). Here I came to see and appreciate your meticulously prepared presentations. My first invited session at a conference was by Erick Delage for a conference in Montreal. I also remember the other moments we met during workshops and conferences in Eindhoven, Tokyo and Houston. Monique Laurent taught the LNMB PhD course which I took during my master studies back in 2012. Here I first encountered advanced methods in nonlinear optimization, something that helped me throughout my three PhD research years. With Kuno Huisman I got insight in the challenges faced when optimization is applied in practice. I learned a lot from the project at ASML to optimize the supplier capacity planning.

During my research I also had the opportunity to go to Technion in Israel to work with Aharon Ben-Tal, a professor I greatly respected for, among others, his seminal work on robust optimization. It has been a great honor to learn from you and closely collaborate on one of the chapters in my thesis during my stay at Technion.

The three years of my PhD research, together with my visits abroad required an substantial amount of planning and funding. I gratefully acknowledge the generous funding received from the Netherlands Organization for Scientific Research (NWO) via the Research Talent grant that fully funded my research. All of the planning was made a lot easier by the help of the secretariat at the Econometrics department and support staff. In particular, I would like to thank Korine Bor, Petra Ligtenberg and Marjoleine de Wit for going above and beyond to help me with the uncommon requests associated with my visits abroad.

I was lucky to be at Tilburg university together with a group of very nice colleagues. I would like to thank Trevor (Jianzhe) for working together with me on one chapter and the things that only you could help me with. Whatever it is, Trevor can fix it: from finding the most adventurous apartments to stay during conferences to Chinese telecom providers abroad. My office mate assignment was also very fortuitous. Together with Ahmadreza I have spent long, pleasant, and at times comic, moments in the office. I would like to thank Krzysztof for the tips on my arrival in Haifa and working with artists. There were many more PhD students with whom I had regular lunches and breaks: Alaa, Amparo, Bas, Elisabeth, Jan, Maria, Mario, Marieke, Marleen, Nick, Olga and Viktoryia. Beside the people that were physically at Tilburg, I would like to thank Bram, a former PhD student, for his help on arrival in Boston in the midst of a blizzard and the enlightening Skype calls we regularly had in these three years.

During my times abroad I met many new friends through academia, sports and social events. Eric and Nathan, thanks for rowing together with me in Haifa in a double, so I could stay in shape while still enjoying a beer with you every now and then. I would also like to thank Jovana, Tabea, Fadi, Andrea (showing us around at the most exciting places ever!), as well as all the others where I have spent many evenings after study with. My stay at MIT was mostly a success because of the wonderful people I met there. In particular, I enjoyed all the discussions and nice evenings with Phebe and Nikita. Miles, I would like to thank you not only for creating the Julia package JuMP, the best and most convenient algebraic modelling language there is, but also for the fun moments we had skiing in Vermont, hiking up Mount Washington and sailing across the English Channel. My stay at MIT got off to a great start by the seafood fiestas organized by Nataly. I will most definitely remember the (many) moments where you and Jacopo got married in Lebanon and Italy, as well as the trip through the countryside with a nice group of other MIT people. Next to intensive research at MIT, social events and food fiestas, I could stay physically and mentally in shape due to the MIT Rowing Club, where I spend in total a year rowing. I would like to thank all my crew members rowing on the Charles river and those joining for the unforgettable ‘convergent’ event in South Korea. Special thanks to crew member and friend Namir, who arranged the frequent social events and defied all odds by passing through all security checks to visit me in Israel. I am also indebted to two nice Dutch friends during my stay at MIT: Maarten from the first trip and Steffie on the second trip. It was good to have some Dutch humor and interaction across the pond.

At home in Europe, I was blessed to have many friendships lasting years or even decades, and there is only space to name a few here. I could stay in shape, physically and mentally, by speed skating with the other students at Braga and friends from the VHL in Utrecht. During the first period, right after starting my PhD, I visited London every now and then to meet with friends and homies: Beatrice, Sarah, Charlotte, Brian, Sophie and April. I am grateful to my friends from my high school period: Efraïm, Imre, Joris, Jorrit, Maarten, Mark, Niels, Olof, Patrick en Pepijn. I would have never made it this far without my friends from the econometrics undergraduate program, and especially Jorn and Kamiel who were always ready to challenge me by asking more questions. Joris and Rick, you transformed (socially and physically) from speed skating buddies to become fantastic friends. The adventures we had, with the trip in the US as a highlight, generated memories that will last.

I would like to thank my aunts, uncles, cousins, grandmothers and grandfathers for their support over a long period of time. Feya, I would like to thank you for being my lovely sister for all this time. Jeroen, thank you for driving me to the ice rink during winter in the Netherlands and all the other support. Of course, I also express gratitude towards my little nephew Twan for his happy support starting from when he was born in the middle of my PhD.

Lotte, your love and support throughout, but especially the last year of my PhD, was fantastic. Despite moments where I seemed absent-minded while finishing the last thoughts in my head, or temporarily moved to Israel, you continued to support me. I cherish the wonderful moments and adventures we had during this time. For this, and for many more you gave, I am deeply thankful.

Tars, my dad, you have been a great inspiration for me and I believe much of my enthusiasm, for research and beyond, I have inherited from you. I am extremely grateful how you have been there for me along this journey. You have travelled to all the places I stayed during my PhD and I have fantastic memories from the trips we made in Boston, Haifa, Houston and the moment you and mum picked me up in London by boat. You have also been there for me in the difficult moments we had before the passing of mum, as well as the time after. You are an amazing dad.

Finally, but most importantly, I would like to thank the person who has always been there for me most, my mother Yolanda. You were there all the time to support my enthusiasm with love: from learning to ride my tricycle, through primary, secondary, undergraduate and graduate studies. All the love and care you gave enables me to freely and confidently exploit all new things I find in life. In my early years, I would go and find out how to ride my tricycle from the ramp near our house. After this accomplishment I would run back towards the house and shout: ‘Look what I did mum!’. Later on, I kept doing this after reaching various milestones in my life, though admittedly not as spectacular as the tricycle achievement. I still remember that in the last weeks we had together I had my first two papers accepted and I could again tell you what I did. While it is difficult for me that I cannot share this celebratory moment of finalizing my thesis in person with you, I can still look up and show you what I could do with all that you gave me. It is therefore to you, mum, that I dedicate this thesis.

Tilburg, The Netherlands
November 2017 Frans de Ruiter

‘Look what I did mum!’

To my mother, Yolanda de Ruiter-Jacobs.

Contents

CHAPTER 1

Introduction

In practice, most decision making problems are based on incomplete or uncertain information. Robust optimization is a methodology to model and solve problems affected by uncertainty in an efficient way. In robust optimization models some decisions have to be made directly (here-and-now) and some decisions can be made later based on extra information that is revealed in the meantime (wait-and-see). To be able to solve models with wait-and-see decisions in robust optimization one has to apply adjustable robust optimization techniques. In this chapter, we first explain robust optimization (Section 1.1) and then adjustable robust optimization (Section 1.2). At the end of this chapter we given an overview of the contributions made in this thesis (Section 1.3).

1.1 Robust optimization

1.1.1 Uncertainty in optimization models

Parameters in optimization models are uncertain due to various reasons. There could be measurement errors if parameter values are obtained via physical experiments. Another reason is that models often include parameter values for which the value is only known in the future, such as future demand realizations or asset returns. Inexact data is another source of uncertainty and is a real issue when inventory records are rounded, historical observations are missing or mistakes were made when entering the data. Even if there is no uncertainty in the input parameters, one might incur some uncertainty when implementing the solution. For example, an optimal design of a physical product cannot be shaped in the exact optimal dimensions specified by the solution.

Robust optimization is one particular way to deal with uncertainty in optimization problems and we motivate and explain it in the rest of this section. Another approach that should be mentioned, but is not considered in this thesis, is stochastic optimization, see Birge and Louveaux (2011). Stochastic optimization approaches rely more on sampling techniques or bounding the expected objective value and probability on constraint violations. In robust optimization one tries to look for solutions that give certain worst-case guarantees on the objective value and feasibility. Robust optimization does not require information on the exact distribution of the uncertain parameter. Furthermore, robust optimization models can, in general, be solved more computationally efficiently than stochastic optimization models. Because of the different objectives and characteristics, the two approaches can be seen as complementary and applicability depends on the underlying motivation of the user. For example, if a decision is repeated often, and incidental high objective values or infeasibilities are not a problem, then a user might prefer to solve the problem with an expected objective value. If, on the other hand, it is a one-time decision or the process is repeated only a few times, the user might want some safe guarantees on the worst-case objective value and insists on more strict feasibility requirements. Finally, we note that in recent years these two fields are starting to converge due to distributionally robust optimization, a research field that started with the paper by Delage and Ye (2010). For more information on distributionally robust optimization we refer the reader to Hanasusanto et al. (2015a).

1.1.2 Illustrative example

Let us illustrate the effect of uncertainty in the parameters on the following toy example:

$$\begin{aligned} \max_{x_1,x_2} \quad & 5x_1 + x_2 \\ \text{s.t.} \quad & 21.94174x_1 + 4.38776x_2 \leq 200 \\ & x_1, x_2 \geq 0. \end{aligned}$$

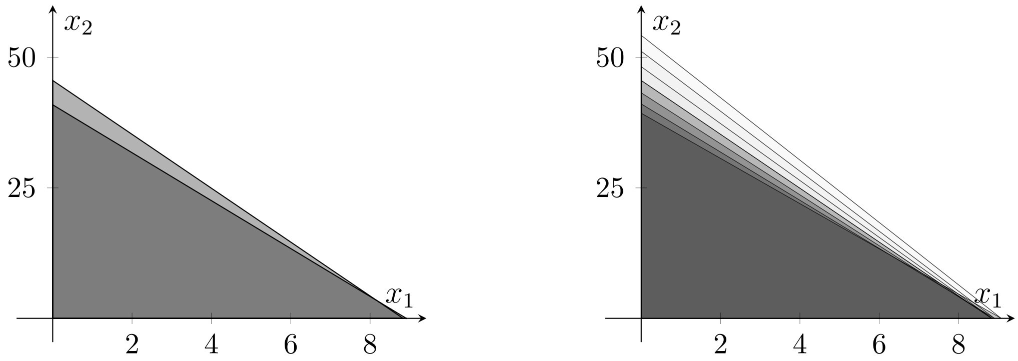
Since this is a very small example, the optimal solution is readily obtained and equal to $x_1 = 0$, $x_2 = 200/4.38776$ which gives an objective value of 45.61. We call this model without uncertainty in the parameters the nominal model and the optimal solution the nominal solution. In practice, coefficients such as 21.94174 and 4.38776 are unlikely to be known up to the precision given, so the real constraint reads as

$$(21.94174 + \zeta_1)\, x_1 + (4.38776 + \zeta_2)x_2 \leq 200, \tag{1.1}$$

with ζ_1 and ζ_2 uncertain parameters. We can see that any $\zeta_2 > 0$ makes the nominal solution infeasible. So for any distribution that is symmetric around $\zeta_2 = 0$ we have that the probability of infeasibility is 50%. To obtain a robust solution, we define an uncertainty set for (ζ_1, ζ_2). As we see later on in Section 1.1.3, there are many choices of uncertainty sets possible, but for now we stick to

$$\mathcal{U} = \{(\zeta_1, \zeta_2) : \ \zeta_1^2 + \zeta_2^2 \leq 0.5\}.$$

Suppose the realization of the uncertain parameter, which we observe after implementing the solution, is $\zeta_1 = 0$, $\zeta_2 = \sqrt{0.5}$. Then for the nominal solution the value on the left-hand side of (1.1) is $(4.38776 + \sqrt{0.5}) \cdot 200/4.38776 = 232.23$, violating the right-hand side with 32.23, or 16%. Hence, the nominal solution is clearly infeasible. If the realization was $\zeta_1 = 0.5$, $\zeta_2 = 0.5$, then the situation would again be completely different, see Figure 1.1a.



(a) Grey area contains solutions satisfying constraint (1.1) for both $(\zeta_1, \zeta_2) = (0.5, 0.5)$ and $(\zeta_1, \zeta_2) = (0, \sqrt{0.5})$. Solutions in the light gray area are only feasible for one realization.

(b) The feasible regions for different values of ζ within $\mathcal{U}$ overlap at some points. The darkest area is the robust feasible region defined by (1.2) containing the solutions that are feasible for all $\zeta \in \mathcal{U}$.

Figure 1.1 – Feasible region for two realizations within the uncertainty set, compared to the robust feasible region.

To obtain a solution that is robust, i.e., feasible regardless of the realization of (ζ_1, ζ_2) within the uncertainty set, we have to solve the following model:

$$\begin{aligned} \max_{x_1, x_2} \quad & x_1 + x_2 \\ \text{s.t.} \quad & (21.94174 + \zeta_1)x_1 + (4.38776 + \zeta_2)x_2 \leq 200 \qquad \forall \zeta \in \mathcal{U} \\ & x_1, x_2 \geq 0. \end{aligned}$$

This model is a semi-infite optimization problem because it has an infinite number of constraints. Fortunately, we can reformulate this constraint as

$$\begin{aligned} (21.94174 + \zeta_1)x_1 + (4.38776 + \zeta_2)x_2 \leq 200 \qquad & \forall \zeta \in \mathcal{U} \\ \Leftrightarrow & \\ 21.94174x_1 + 4.38776x_2 + \max_{\zeta \in \mathcal{U}} \{\zeta_1 x_1 + \zeta_2 x_2\} \leq 200 & \\ \Leftrightarrow & \\ 21.94174x_1 + 4.38776x_2 + \tfrac{1}{\sqrt{2}}\sqrt{x_1^2 + x_2^2} \leq 200. & \end{aligned} \tag{1.2}$$

The last line (1.2) is a second-order cone constraint. The feasible region that is formed by this second-order cone constraint is depicted in Figure 1.1b. The optimal solution to the robust model is $x_1 = 8.48764$ and $x_2 = 1.74231$ leading to an objective value of 44.18. This robust objective value is only 3.2% lower than the nominal objective value, but does protect for all realizations within the uncertainty set (1.1).

The toy example given in this section is a very small example. Ben-Tal and Nemirovski (2002) show that for some much larger problems from the NETLIB library constraint violations can be very severe if uncertainties are neglected, whereas the robust solution has only a slightly lower (assuming we are maximizing) optimal objective value than the optimal nominal objective value.

1.1.3 Robust counterparts

In the illustrative example from the previous section we determined a solution that is a priori known to be feasible for any realization of the uncertain parameter within the uncertainty set. This illustrates the fundamental principles underlying the robust optimization paradigm:

1. All decisions are here-and-now and have to be made before the realizations of the uncertain parameters are known.
2. The uncertain parameter resides in a prespecified uncertainty set $\mathcal{U}$.
3. All constraints are "hard", i.e., no violations are allowed for any realization of the uncertain parameter in the uncertainty set.

There are several methods that extend the scope of robust optimization by relaxing the conditions in the above principles. One example is distributionally robust optimization which adapts the second principle. Those techniques assume that crude probabilistic information such as the mean and the variance of the uncertain parameter are known, see Delage and Ye (2010). There are also methods that relax the third principle, such as comprehensive or globalized robust counterparts (Ben-Tal et al. 2006; Ben-Tal et al. 2017) or light robustness (Fischetti and Monaci 2009). If we relax the first principle, then we allow some decisions to be made after the realization of the uncertain parameter is known. This is called adjustable robust optimization, which will be explained in Section 1.2.

The robust counterpart constraint (1.2) from our illustrative example is called the tractable robust counterpart. Let us now consider a general linear robust model of the form

$$
\begin{aligned}
\min_{x} \quad & c^\top x \\
\text{s.t.} \quad & \left(a^i + A^i\zeta\right)^\top x \le r_i \qquad i = 1, \ldots, m, \ \forall \zeta \in \mathcal{U},
\end{aligned}
\tag{1.3}
$$

where $x \in \mathbb{R}^{n_x}$ are the here-and-now decisions and $c \in \mathbb{R}^{n_x}$ the objective coefficients. The uncertain parameter ζ resides in a compact convex uncertainty set $\mathcal{U} \subset \mathbb{R}^{n_\zeta}$. There are m constraints with coefficients given by $(a^i + A^i\zeta)$, with $a^i \in \mathbb{R}^{n_x}$ the nominal value, $A^i \in \mathbb{R}^{n_x \times n_\zeta}$ and right-hand side $r_i \in \mathbb{R}$, for all $i = 1, \ldots, m$. Before we show how tractable robust counterparts are formulated, we make a few remarks regarding this model and its generality:

- Although there is a common parameter ζ affecting all the constraints simultaneously, we can without loss of generality consider the uncertainty constraint-wise. That is, the following two sets of uncertain constraints are equivalent:

$$\forall \zeta \in \mathcal{U}: \begin{cases} \left(a^1 + A^1\zeta\right)^\top x \leq r_1 \\ \left(a^2 + A^2\zeta\right)^\top x \leq r_2 \\ \vdots \\ \left(a^m + A^m\zeta\right)^\top x \leq r_m \end{cases} \Leftrightarrow \begin{cases} \left(a^1 + A^1\zeta\right)^\top x \leq r_1 & \forall \zeta \in \mathcal{U} \\ \left(a^2 + A^2\zeta\right)^\top x \leq r_2 & \forall \zeta \in \mathcal{U} \\ \vdots \\ \left(a^m + A^m\zeta\right)^\top x \leq r_m & \forall \zeta \in \mathcal{U}. \end{cases}$$

  The set of uncertain constraints on the right seem more restrictive as for each constraint we could pick a different $\zeta \in \mathcal{U}$. However, suppose for some candidate solution x there exists an $\zeta \in \mathcal{U}$ that violates the i-th constraint in the system on the right. Then for the same ζ the uncertain constraint in the left set of constraints is violated.

- Uncertainty in the objective function, e.g., an objective function $\left(a^0 + A^0\zeta\right)^\top x$ can be included by introducing a new variable z and adding the constraint $\left(a^0 + A^0\zeta\right)^\top x \leq z$. The value of z is our new objective which does not contain uncertain parameters and therefore fits the format of (1.3).

- In model (1.3) there is no uncertainty in the right-hand side. This is deliberately done for ease of exposition, but we can incorporate an uncertain right-hand side by introducing an extra variable and enforce it equal to 1, i.e., by adding the constraint $x_{n_x+1} = 1$.

By the first remark above it becomes clear that we can consider each constraint $i = 1, \ldots, m$

$$\left(a^i + A^i\zeta\right)^\top x \leq r_i \qquad \forall \zeta \in \mathcal{U}, \tag{1.4}$$

separately for reformulation purposes. Furthermore, the constraint (1.4) is satisfied for all values of $\zeta \in \mathcal{U}$ if and only if it is satisfied for the value of $\zeta \in \mathcal{U}$ that maximizes the value on the left-hand side of the constraint:

$$\left(a^i + A^i\zeta\right)^\top x \leq r_i \qquad \forall \zeta \in \mathcal{U} \qquad \Leftrightarrow \qquad \max_{\zeta \in \mathcal{U}} \left\{ \left(a^i + A^i\zeta\right)^\top x \right\} \leq r_i. \tag{1.5}$$

The final formulation of the tractable robust counterpart is obtained using duality for linear optimization (in case $\mathcal{U}$ is a polyhedral set) or duality for convex optimization (for general convex sets $\mathcal{U}$). In the next example, we show how to obtain the tractable robust counterpart for a polyhedral uncertainty set. This set is also used to describe uncertainty in Chapters 2 and 3.

Example 1.1 *Let* $\mathcal{U} = \{\zeta \geq 0 : \ D\zeta \leq d\}$*, where* $D \in \mathbb{R}^{p \times n_\zeta}$ *and* $d \in \mathbb{R}^p$ *for some integer* p *such that* $\mathcal{U}$ *is nonempty. Given* x*,* (1.5) *can be reformulated as follows:*

$$\max_{\zeta \in \mathcal{U}} \left\{ \left(a^i + A^i \zeta\right)^\top x \right\} \leq r_i$$

$$\Leftrightarrow$$

$$\left(a^i\right)^\top x + \min_{\lambda \geq 0} \left\{ d^\top \lambda : \ D^\top \lambda \geq \left(A^i\right)^\top x \right\} \leq r_i$$

$$\Leftrightarrow$$

$$\begin{cases} (a^i)^\top x + d^\top \lambda \leq r_i \\ D^\top \lambda \geq (A^i)^\top x \\ \lambda \geq 0. \end{cases}$$

In the second line we have used strong duality for linear optimization. In the last line we used the fact that the constraint is satisfied for the minimizer $\lambda \geq 0$ *if and only if there exists a* $\lambda \geq 0$ *that satisfies both* $(a^i)^\top x + d^\top \lambda \leq r_i$ *and* $D^\top \lambda \geq (A^i)^\top x$*. Note that in the final statement* λ *does not necessarily has to be the minimizer.*

Some other uncertainty sets that are used often are given in Table 1.1 and many more can be found in Ben-Tal et al. (2015).

Table 1.1 – Examples of uncertainty sets and their robust counterparts. The parameter Γ controls the size of the uncertainty set and $\|\cdot\|_p$ is the p-norm.

Uncertainty set	$\mathcal{U}$	Robust counterpart of $(a^i + A^i\zeta)^\top x \leq r_i \quad \forall \zeta \in \mathcal{U}$
Box	$\{\zeta : \|\zeta\|_\infty \leq \Gamma\}$	$(a^i)^\top x + \Gamma \left\|(A^i)^\top x\right\|_1 \leq r_i$
Ball	$\{\zeta : \|\zeta\|_2 \leq \Gamma\}$	$(a^i)^\top x + \Gamma \left\|(A^i)^\top x\right\|_2 \leq r_i$
Budget	$\{\zeta : \|\zeta\|_\infty \leq 1, \ \|\zeta\|_1 \leq \Gamma\}$	$(a^i)^\top x + \left\|(A^i)^\top x - \lambda\right\|_1 + \Gamma \|\lambda\|_\infty \leq r_i$
Polyhedral	$\{\zeta \geq 0 : \ D\zeta \leq d\}$	$\begin{cases} (a^i)^\top x + d^\top \lambda \leq r_i \\ D^\top \lambda \geq (A^i)^\top x \\ \lambda \geq 0 \end{cases}$

If the uncertainty set is a box, then the resulting tractable robust counterpart contains a 1-norm, which can be written as a compact linear optimization model using additional variables. The term containing the 1-norm in the tractable robust counterpart constraint can be seen as an extra safeguard for the nominal constraint $(a^i)^\top x \leq r_i$, which itself depends on the decision x. For the other uncertainty sets there are similar safeguards, all depending on the decision x. If the uncertainty set is a ball, then the tractable robust counterpart contains a 2-norm term as a safeguard, making the constraint a second-order cone constraint. All of these models can be solved efficiently in theory and practice with modern solvers, even when the models contain thousands of variables and constraints. Throughout this section we focussed on linear constraints for ease of exposition. We can also consider nonlinear constraints of the form $f(\zeta, x) \leq 0$ that are concave in the uncertain parameter ζ for every x and convex in x for every ζ. These techniques rely on duality for convex optimization such as Fenchel duality, see Ben-Tal et al. (2015). For more general information and an overview of applications of robust optimization, we refer to the book by Ben-Tal et al. (2009) and the surveys by Bertsimas et al. (2011b) and Gabrel et al. (2014b).

1.1.4 Origins of Robust Optimization

One of the early appearances of robust optimization techniques was in Soyster (1973), where the authors considered box uncertainty sets. That paper introduced the same reasoning and even much of the notation as used in robust optimization today. Surprisingly, not much happened with that paper in the operations research community until the end of the 90s as can be seen in Figure 1.2. The first series of papers that

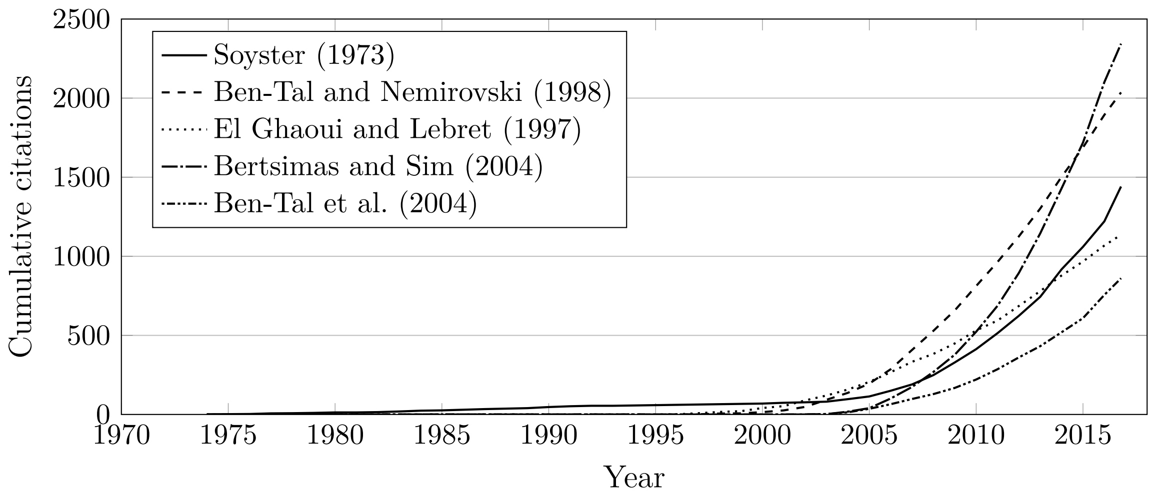


Figure 1.2 – Cumulative citations for some of the influential papers on Robust Optimization. Research only really started to take of at the end of the '90s (data from Google Scholar).

established robust optimization as a field are by Ben-Tal and Nemirovski (1998), Ben-Tal and Nemirovski (1999), El Ghaoui and Lebret (1997) and El Ghaoui et al. (1998). A decade after the first papers, the (up to now only) book on robust optimization by Ben-Tal et al. (2009) was published, detailing many of the techniques used to formulate tractable robust counterparts and a lot of applications. A few years after the first seminal papers, the paper by Bertsimas and Sim (2004) appeared which described the polyhedral budget uncertainty. That paper received a lot of attention and is as of today the most cited paper in robust optimization according to Google Scholar. Two major active subfields in robust optimization are adjustable robust optimization, introduced in the next section, and distributionally robust optimization, for which we refer to Hanasusanto et al. (2015a).

1.2 Adjustable robust optimization

1.2.1 Wait-and-see decisions in optimization models

Robust optimization as described in Section 1.1 only deals with *here-and-now* variables: the decisions have to be made before any information on the uncertain parameter is known. Decision making problems often contain some decisions of a *wait-and-see* type, which can be decided upon whenever (part of) the realization of the uncertain parameter is known. Wait-and-see decisions arise naturally in many multistage optimization applications where decisions are made in different time periods. We list a few applications, that appear in this thesis, below and explain what the here-and-now decisions, wait-and-see decisions and the uncertain parameters are.

Facility location planning. (Sections 2.6 and 5.2.3)
The here-and-now decisions determine which distribution centers are opened. The actual distribution plan only has to be made after the uncertain demand from each customer is known. Hence, the transportation quantities from the facilities to the customers are the wait-and-see decisions.

Inventory management. (Sections 5.2.1 and 6.4)
In each period, the order quantities are placed after the uncertain demand from previous period is observed. The here-and-now decisions are therefore at the beginning of the planning horizon and the wait-and-see decisions from the second period onwards.

Lot-sizing and distribution on a network. (Sections 2.5 and 3.4)
If we have an inventory model with multiple warehouses at different locations, then stock allocation decisions are important as some warehouses may be closer to (or further away from) some customers and have different storage costs. The here-and-now decisions in this application are the stock sizes at each warehouse. Demand at the customer locations are the uncertain parameters. The wait-and-see decisions are

the transportation quantities from the warehouses to the customers.

Wireless sensor networks. (Section 3.5)
In the wireless sensor location problem there is a set of sensors in a field or at sea, whose locations are subject to uncertainty due to e.g., drift at sea or due to inexact placements via air drops. We have to install some of the (interconnected) transmission modules here-and-now before the exact location of the sensors is known and for some modules we could perhaps wait-and-see until the precise locations of the sensors is known.

There are many more applications of adjustable robust optimization to multistage problems beside the ones that appear in this thesis. Examples include: management of power systems (Bertsimas et al. 2013; Ng and Sy 2014), project management (Wiesemann et al. 2012), portfolio optimization (Calafiore 2008; Calafiore 2009; Rocha and Kuhn 2012), dynamic pricing (Adida and Perakis 2006) and capacity expansion planning (Ordóñez and Zhao 2007).

The second, perhaps less obvious, way in which wait-and-see decisions arise in robust optimization models is because of the use of auxiliary (or analysis) variables. In many model formulations auxiliary variables are used to evaluate parts of the objective value such as the backlog or holding costs in each period. These variables are required to formulate the model as a tractable linear or convex optimization model. Examples are robust sum-of-max problems, such as inventory models, which are nonconvex but can be modeled as a convex problem with the use of auxiliary variables. An illustrative example of this type of problems is given in Section 1.2.3, as well as the wireless sensor network in Section 3.5. Contrary to the (here-and-now) primary decisions, auxiliary variables are always allowed to have a wait-and-see character. Their values never have to be implemented because they are only used to evaluate the cost of the solution. Restricting these auxiliary variable to be here-and-now can be severely conservative as shown in Gorissen and den Hertog (2013), Gorissen et al. (2015), Delage and Iancu (2015), Ardestani-Jaafari and Delage (2016a), and Ardestani-Jaafari and Delage (2016b). The use of auxiliary variables is also illustrated in the example in Section 1.2.3.

1.2.2 Model formulation and linear decision rule solutions

Adjustable robust optimization is a way to extend the static robust model (1.3) to account for wait-and-see decisions. The formulation of the constraints is such that for all ζ in the uncertainty set $\mathcal{U}$ there exists a wait-and-see decision $y \in \mathbb{R}^{n_y}$ that satisfies the constraints:

$$
\begin{aligned}
&\max_{x} \quad c^\top x \\
&\text{s.t.} \quad \forall \zeta \in \mathcal{U}\ \exists y \in \mathbb{R}^{n_y} : \left(a^i + A^i\zeta\right)^\top x + \left(b^i\right)^\top y \le r_i \qquad i = 1, \dots, m,
\end{aligned}
\tag{1.6}
$$

where c, a^i, A^i and r_i are as in (1.3) and we have the additional coefficients $b^i \in \mathbb{R}^{n_y}$ for the wait-and-see decisions y. We make several remarks regarding this model formulation:

- As in the static robust model, uncertainty in the objective function, e.g., an objective function $(a^0 + A^0\zeta)^\top x + (b^0)^\top y$ can be included by introducing a new here-and-now variable z. We then have to include the constraint $(a^0 + A^0\zeta)^\top x + (b^0)^\top y \le z$. The value of z is our new objective which does not contain uncertain parameters itself.

- We consider the *fixed recourse* case, i.e., the value of b^i does not depend on the uncertain parameter ζ.

Contrary to model (1.3), the adjustable robust model (1.6) is in general difficult to solve. In fact, it has been shown in Guslitzer (2002) that this model is NP-hard to solve for polyhedral uncertainty sets. Intuitively, the reason is that the wait-and-see decision y can be seen as a function: for every realization ζ one has to choose a different value for the wait-and-see decision y. Therefore, to find the optimal solution, one has to optimize over functions. This makes the problem an infinite-dimensional optimization problem. This class of optimization problems is notoriously difficult to solve to optimality. Fortunately, there are computationally efficient methods that give very good solutions. One of the most popular and versatile methods nowadays is based on linear decision rules (also called affine policies or affine control), which was introduced in the seminal paper on adjustable robust optimization by Ben-Tal et al. (2004). Instead of allowing the wait-and-see decision to be any function, we restrict the possible class of functions to be affine:

$$
y(\zeta) = u + V\zeta, \tag{1.7}
$$

where $u \in \mathbb{R}^{n_y}$ and $V \in \mathbb{R}^{n_y \times n_\zeta}$ is a vector and a matrix. Each entry in u or V is a new here-and-now decision and together they determine the affine dependence on ζ. Substituting the decision rule (1.7) in model (1.6) we obtain

$$
\begin{aligned}
&\max_{x,u,V} \quad c^\top x \\
&\text{s.t.} \quad \forall \zeta \in \mathcal{U} : \left(a^i + A^i\zeta\right)^\top x + \left(b^i\right)^\top (u + V\zeta) \le r_i \qquad i = 1, \dots, m.
\end{aligned}
\tag{1.8}
$$

Model (1.8) is again a normal linear robust optimization model, which also fits the format of model (1.3). Therefore, its tractable robust counterpart can be formulated using any of the methods described in Section 1.1. Since we restricted ourselves to linear decision rules, model (1.8) is called the affine adjustable robust counterpart model. These linear decision rules can be used for two-stage problems, but also for multistage optimization problems. In multistage settings we have to ensure that decision rules for the wait-and-see decisions are nonanticipative, i.e., they do not use information that is not available at the time that the wait-and-see decision has to implemented. For linear decision rules this can be enforced by setting some of the elements in the matrix of variables V to zero. If the k-th wait-and-see decisions can only use information up to stage $k-1$, then $V_{k,j}$ (the variable in the k-th row and j-th column) for $j = k+1, \dots, n_\zeta$ is set to zero in (1.7).

1.2.3 A small inventory management example

Consider an inventory model with just one product and uncertain demand over two weeks.[1] The current inventory level is $I_0 = 5$ and at the beginning of each week we can place orders to replenish our inventory. We denote the order quantities in the first and second week by respectively q_1 and q_2 units. The supplier that delivers the demand has informed us that in the second week he can deliver at most $q_2^{max} = 3$ units. If the inventory level is positive, then holding costs of $h = 1$ euro per unit are incurred and if the inventory is negative, we incur backlog costs of $b = 2$ euro per unit. When the demand d_1 an d_2 is certain, we can model this as a nominal optimization model as follows:

$$\min_{q_1,q_2} \sum_{t=1}^{2} \max\left\{ h\left(I_0 + \sum_{s=1}^{t} (q_t - d_t)\right),\ -b\left(I_0 + \sum_{s=1}^{t} (q_t - d_t)\right)\right\},$$

where the sum-of-max ensures that in case the inventory level is positive holding cost are incurred and in case of negative inventory level we have backlog costs. The sum-of-max also makes this a nonlinear optimization problem. With the introduction of auxiliary variables we can write this as a linear model, see (1.9), where the auxiliary variables c_1 and c_2 represent the costs for the first and second week respectively. Now consider the case with uncertain demand: $d_t = 5 + \zeta_t$, where for the uncertainty set we take $\{(d_1, d_2) :\ d_t = 5 + \zeta_t,\ t = 1, 2,\ \|(\zeta_1, \zeta_2)\| \le 5\}$, i.e., the demand lies in a ball centered at the nominal demand $(5, 5)$ with radius 5. The order quantity in the first period is a here-and-now decision because it has to be made before any demand is observed. The wait-and-see decisions in this example are:

- The order quantity in the second week, q_2. This can be made after we observe the demand d_1 from the first week.

[1] This model is an adapted version of the one in Gorissen et al. (2015, Section 10)

- The auxiliary variables c_1 and c_2. These are auxiliary variables to evaluate the holding/backlog costs in each period and are allowed to be chosen after d_1 and d_2 are known.

$$
\begin{aligned}
\min_{q_t,c_t} \quad & \sum_{t=1}^{2} c_t \\
\text{s.t.} \quad & c_t \geq h\left(I_0 + \sum_{s=1}^{t}(q_t - d_t)\right) && t = 1,2 \\
& c_t \geq -b\left(I_0 + \sum_{s=1}^{t}(q_t - d_t)\right) && t = 1,2 \\
& q_t \geq 0 && t = 1,2 \\
& q_2 \leq q_2^{max}.
\end{aligned}
\tag{1.9}
$$

We solve the adjustable robust version of (1.9) with linear decision rules. For the order quantity in the first week the linear decision rule becomes

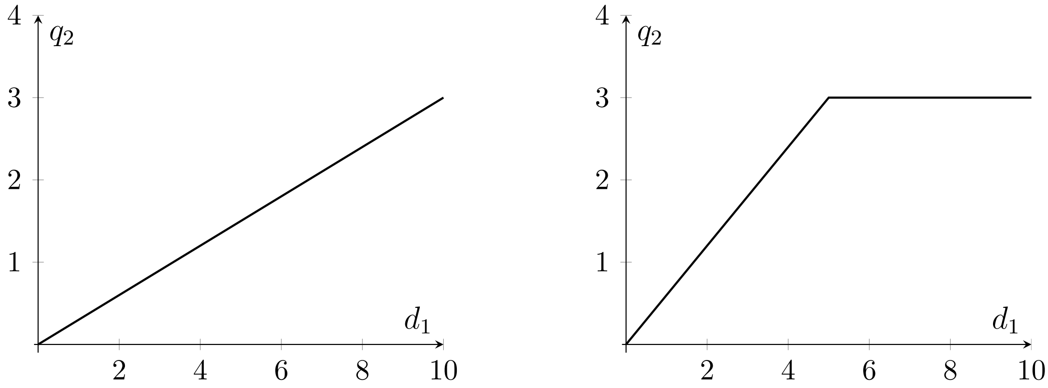
$$q_2(d_1) = \bar{q}_{2,0} + \bar{q}_{2,1}d_1,$$

where $\bar{q}_{2,0}$ and $\bar{q}_{2,1}$ are new variables that have to be decided here-and-now. We can make a similar linear decision rule for c_t:

$$c_t(d_1, d_2) = \bar{c}_{t,0} + \bar{c}_{t,1}d_1 + \bar{c}_{t,2}d_2,$$

with $\bar{c}_{t,0}$, $\bar{c}_{t,1}$ and $\bar{c}_{t,2}$ here-and-now variables for $t = 1,2$. Note that q_2 only depends on d_1 (nonanticipative), whereas c_t depend on both d_1 and d_2 for $t = 1,2$. The affine adjustable robust variant of (1.9) can be reformulated using the robust counterpart formulation corresponding to the ball uncertainty set in Table 1.1. We programmed this example in Julia using the JuMP optimization package developed by Dunning et al. (2017) and the commercial solver Mosek 8 (any conic solver can solve this small model).[2] We display the solution of the linear decision rule in Figure 1.3a. We also display another decision rule in Figure 1.3b. This is a decision rule of the form $q_2(d) = \bar{q}_{2,0} + \bar{q}_{2,1}d_1 + \hat{q}_{2,0}|d_1 - 5|$, which has an additional absolute value term (also for the costs c_1 and c_2). These nonlinear decision rules are introduced in Chapter 4. The optimal objective value with linear decision rules is 14.78 euros, which can be improved by using the nonlinear decision rule to 13.67 euros. Note that in both cases we take the auxiliary variables to be wait-and-see variables as argued in Section 1.2.1. If we would have chosen the auxiliary variables c_1 and c_2 to be here-and-now, then the optimal objective value, with linear decision rule for q_2 would be much higher, namely 18.67 euros.

[2]Code is available online at www.fransderuiter.com.



(a) Linear decision rule $q_2 = 0.3d_1$.
Optimal here-and-now: $q_1 = 4.11$.
Objective value: 14.78 euros.

(b) Nonlinear decision rule:
$q_2 = 1.5 + 0.3d_1 - 0.3|d_1 - 5|$.
Optimal here-and-now: $q_1 = 3.50$.
Objective value of 13.67 euros.

Figure 1.3 – Two different solutions of here-and-now q_1 and wait-and-see q_2 for model (1.9)

1.2.4 Other solution methods

There are many more methods than the linear decision rules method from Section 1.2.2. Here we list some of the most prominent alternatives in the literature.

Folding horizon approaches. In multistage optimization approaches the most important decision is the here-and-now decision. In inventory problems, as well as many more applications, we can implement and fix the here-and-now decision, observe the realization of the uncertain parameter and then re-optimize for the remaining stages. We can do this for each stage, "folding the model" up to the end of the planning horizon. Notice that one must provide a feasible here-and-now decision to be implemented in the first stage. This decision has to be found using some other technique, for example by using linear decision rules. It can therefore also be seen as a complementary method instead of an alternative to linear decision rules. We apply this procedure in Chapter 5. This approach is also known as the receding or shrinking horizon approach, see Delage and Iancu (2015).

Sampling. This method considers a finite subset of N scenarios from the uncertainty set. This sampling method was first considered by Calafiore and Campi (2005) for the static robust models and later for adjustable robust optimization models in Bertsimas and Caramanis (2007). The sampled version of (1.6) is just a regular (not even an uncertain) small linear optimization model. Unfortunately, this method is not really a solution method since it only guarantees feasibility for a subset of scenarios from the uncertainty set. The here-and-now decisions found with this method could be infeasible and the objective value only provides a lower bound to the optimal objective value of the adjustable robust optimization model. Nevertheless, it is still a useful approach to measure the quality of other solution methods. If the subset is chosen carefully, it can give strong lower bounds with only a small set of scenarios. One way to choose the set of scenarios is to take the set of scenarios which is binding for the affine adjustable robust model. This method was introduced in Hadjiyiannis et al. (2011) and is further improved in Chapters 2 and 3.

Benders decomposition. By duality it can be shown that adjustable robust optimization is equivalent to a bilinear optimization model. These models can be solved using Benders style decomposition algorithms and several papers describe variations of these algorithms, such as Thiele et al. (2009), Zhao and Zeng (2012), Zeng and Zhao (2013), Bertsimas et al. (2013), and Gabrel et al. (2014a). An interesting note is that this single dualization step to obtain the bilinear optimization model was what initially inspired the research in Chapter 2 to further dualize the bilinear optimization model.

Finite adaptability. Rather than having a continuous decision rule, we could also restrict the wait-and-see decision to take a finite number of values. This finite adaptability approach was introduced by Bertsimas and Caramanis (2010) and later studied in Hanasusanto et al. (2015b). In practice, these approaches have the benefit that the end user is faced with a finite set of possible actions to prepare for. Another advantage of this approach is that it can include both continuous and integer valued wait-and-see decisions. The difficulty is that, regardless of the continuity of the wait-and-see decisions, the resulting model is often a quite large mixed integer optimization model.

Partitioning the uncertainty set. Closely related to finite adaptability are methods that partition the uncertainty set in K different sets and take a different here-and-now decision, or decision rule, for each partition. This was first done by Ben-Ameur (2007) and later in more general setting by Vayanos et al. (2011). In those papers the partition was made a-priori. An algorithmic approach that refines the partition in each iteration was recently introduced by Postek and den Hertog (2016) and Bertsimas and Dunning (2016). The benefit of these approaches is that they can also be applied if the wait-and-see decisions are integer. The papers include some examples where the method performs very well. However, improvement is not guaranteed in each refinement, so the solution might not converge for all problems. Furthermore, the model size grows as the partition is refined further.

Fourier-Motzkin elimination. Recently it has been shown by Zhen et al. (2016) how Fourier-Motzkin elimination can be used to efficiently eliminate wait-and-see decisions from adjustable robust optimization models. Although Fourier-Motzkin elimination yields an exponential number of constraints, when coupled with a smart way of detecting redundant constraints this explosion of the number of constraints is limited. In this way, adjustable robust models of small size can be solved to (near) optimality. For larger models one could eliminate just a few wait-and-see decisions and apply linear decision rules for the remaining variables. This procedure is also used in Section 3.4.

1.2.5 Origins and current research challenges

Adjustable robust optimization only started about 15 years ago with the thesis by Guslitzer (2002) and the paper based on that thesis by Ben-Tal et al. (2004). Interest in multistage optimization models dates back much further to the beginnings of the operations research field. George Dantzig, the founding father of linear optimization, said in 1991:

> "It is interesting to note that the original problem that started my research is still outstanding – namely the problem of planning or scheduling dynamically over time, particularly planning dynamically under uncertainty. If such a problem could be successfully solved it could eventually through better planning contribute to the well-being and stability of the world." (Dantzig 1991, p.30)

More than two decades after Dantzig's statement it is fair to say that this problem is still outstanding today, although a lot of progress has been made. New research developments, combined with incredible computing power nowadays, allows us to solve more challenging models in smart, tractable ways. Adjustable robust optimization is just one way to model and solve dynamic problems under uncertainty. In fact, the "original problem" that George Dantzig mentions above refers to a problem in one of his early papers called *Linear programming under uncertainty* (Dantzig 1955). That paper introduces two-stage stochastic optimization models. Stochastic optimization is a huge research field and very well applicable if probabilistic information is available. However, Dantzig (1991, p.21) also writes that another important criterion of models is that they are computable in a practical way. Adjustable robust optimization approaches and linear decision rules, introduced in the seminal paper by Ben-Tal et al. (2004), aim for exactly that: to be tractable from a theoretical complexity and a practical computational point of view.[3] Linear decision rules have been around for a long time. They appeared in Charnes et al. (1958) where they were used for two-stage stochastic optimization models. Linear decision rules have been used in other communites as well. An early use was in control theory in the

[3]Aharon Ben-Tal once described this to me as "practable" approaches.

thesis by Wisenhausen (1966). Interest in linear decision rules disappeared over time, but was revived by the seminal paper on adjustable robust optimization. There has been many applications and new developments in solution techniques for adjustable robust optimization since 2004, as described in Section 1.2.1 and 1.2.4.

1.2.6 Some remaining challenges in adjustable robust optimization

There are several important challenges remaining in adjustable robust optimization. Some of these are (partially) addressed in this thesis and described below.

Efficiency of solution methods. The affine adjustable robust counterpart model is a convex optimization model which can directly be given to off-the-shelve solvers. However, the size of these models is much larger than the static robust version because the linear decision rules add many additional here-and-now decision variables. As the computational time depends on the size of the model, this makes the models more difficult to solve. Chapter 2 describes a dual approach for general two-stage adaptive linear optimization models. The new dualized model is again a two-stage adaptive linear optimization model, but differs in the number of variables and number of constraints. We show that, for certain problems, the dualized formulation can be solved much faster with linear decision rules than the original primal formulation.

Lower bounds on the optimal value. There are some special problem structures where one can prove that linear decision rules are optimal (Bertsimas et al. 2010; Iancu et al. 2013; Ardestani-Jaafari and Delage 2016b; Gounaris et al. 2013). These proofs give insight in the impressive power of the seemingly restrictive linear decision rules. However, virtually all the numerical examples in the literature do not fit into the special structures described in those papers. Good lower bounds to assess the quality of instance-specific solutions are still required. Some work that partially addresses this challenge is by Kuhn et al. (2011) as well as the sampling method from Hadjiyiannis et al. (2011) that was mentioned in Section 1.2.4. In Chapter 2 we improve the lower bound method from Hadjiyiannis et al. (2011) by including information obtained from the solution of a dualized formulation.

Nonlinear adjustable robust optimization. If models are nonlinear in the wait-and-see decisions, then linear decision rules do not result in tractable convex optimization models. The inability to solve nonlinear adjustable robust optimization models is in sharp contrast with static robust optimization which is as tractable for nonlinear convex optimization problems as it is for linear problems. In Chapter 3 we extend the dual approach from Chapter 2 to nonlinear robust optimization models that are convex, rather than just linear, in the wait-and-see decisions. We show that for general polyhedral uncertainty sets the dualized formulation is linear in the wait-and-see decisions so that linear decision rules, or any other method from Section

1.2.4, can be used to find solutions.

Efficient nonlinear decision rules. Instead of restricting ourselves to linear decision rules we could consider richer classes of nonlinear decision rules. In most cases the model with nonlinear decision rules becomes harder to solve, or can only be solved approximately. Quadratic decision rules are discussed in the book by Ben-Tal et al. (2009) and higher order polynomials by Bertsimas et al. (2011a). The resulting tractable robust counterpart models in those papers can be solved (approximately) by semidefinite models, instead of second-order cone models as with linear decision rules. To be able to scale in the way one can with linear decision rules, nonlinear decision rules methods should require less complex model formulations. In Chapter 4 we introduce some nonlinear decision rules for which the tractable robust counterpart formulation is of the same optimization class as the affine adjustable robust counterpart.

Conservativeness of solutions. One criticism on (adjustable) robust optimization is that the solutions are too conservative because it focusses on worst-case protection. Iancu and Trichakis (2013) show that robust optimization models can have multiple robustly optimal solutions. Among those there can be solutions that give lower costs for each realization within the uncertainty set, i.e., they Pareto dominate the other solutions. Ideally, one tries to find solutions that perform good on other metrics, besides worst-case guarantees, such as the average objective value. In Chapter 5 we remedy the conservativeness of adjustable robust optimization by providing a two step procedure to efficiently choose a solution, among all optimal solutions, that performs best on a secondary requirement such as the average objective value under some distribution.

Inexact data. In adjustable robust optimization, the decision in each stage is a function of the data on the realizations of the uncertain demand gathered from the previous stages. There is much evidence in the information management literature that data quality in inventory systems is often poor. Reliance on data “as is” may then lead to poor performance of “data-driven” methods such as adjustable robust optimization. Chapter 6 describes approaches for cases where the revealed data in each stage of a multistage model is still inexact to some extent.

There are also other important challenges in adjustable robust optimization that are not addressed in this thesis.

Nonfixed recourse. We focus in this thesis on the fixed-recourse situation, where the parameter b^i in (1.6) does not depend on the uncertain parameter. Although there are some methods that can deal with nonfixed recourse models, such as the partitioning methods, much research still has to be done on developing methods and analyzing the performance of methods in the nonfixed recourse case.

Integer wait-and-see decisions. The finite adaptability and partitioning methods can deal with integer wait-and-see decisions. These methods have been proven on small to moderate sized problems and are very promising. The methods are still much more computationally demanding than integer static robust (or nominal) optimization models. An important step forward would be new computationally efficient methods, or to make the existing finite adaptability and partitioning methods more efficient for larger scale problems.

Design of uncertainty sets and learning. Motivation for construction of (data-driven) uncertainty sets for static robust optimization was given in papers such as Ben-Tal et al. (2013) and Bertsimas et al. (2017a). In multistage settings there are extra difficulties such as the effect that decisions have on realizations of uncertain parameters. In many cases, e.g., pharmaceutical drug testing, price-demand curves or finding the best sports team in competitions, the outcomes depend on the decisions made in previous periods. To incorporate learning effects, one could take uncertainty sets $\mathcal{U}(x)$, where the uncertainty set depends on the decision variable x. When x is integer, there have been some results for specific cases such as Vayanos et al. (2011) and Poss (2014). The resulting models in those papers are often large scale MIPS and the performance on practical cases such as drug testing are still unclear. New methods that can efficiently deal with such decision-dependent uncertainty sets would be an important step forward.

1.3 Contributions and outline

In all chapters we adhere to the philosophy behind (adjustable) robust optimization: the resulting model should be computationally tractable in theory and practice. The first two chapters introduce dual approaches to adjustable robust optimization and the last three chapters improve the linear decision rule solutions by using the original (primal) formulation. Below we summarize the contributions per chapter.

In Chapter 2, we come up with a dual approach to linear two-stage adjustable robust optimization models. We show that the optimal primal affine policy can be directly obtained from the optimal affine policy in the dual formulation. We provide empirical evidence that the dualized model, in the context of lot-sizing with distribution on a network and two-stage facility location problems, solves an order of magnitude faster than the primal formulation with affine policies. Furthermore, the affine policy of the dual formulations can be used to provide stronger lower bounds on the optimality of affine policies.

Chapter 3 extends the dual approach from Chapter 2 to nonlinear robust optimization models that are convex in the wait-and-see decisions. We show that the resulting dualized formulation is linear in the wait-and-see decisions so that all methods from Section 1.2.4 can be applied again. We also explain how some static nonconvex optimization models can be modeled in two-stage robust formats using auxiliary variables. We use two numerical examples to illustrate the effectiveness of the dualized formulation. Finally, we show how to obtain lower bounds on the optimal value of the nonlinear two-stage robust optimization model.

Chapter 4 introduces nonlinear decision rules for ellipsoidal and general convex uncertainty sets. The resulting tractable robust counterpart of a model with our nonlinear decision rule is again a convex optimization model of the same optimization class as the original model with linear decision rules. We show both theoretically and via two numerical examples taken from the literature that the new nonlinear decision rules improve over linear decision rules.

Chapter 5 shows that multiple solutions exist for the production-inventory example in the seminal paper on adjustable robust optimization in Ben-Tal et al. (2004). All these optimal robust solutions have the same worst-case objective value, but the mean objective values differ up to 21.9% and for individual realizations this difference can be up to 59.4%. We show via additional experiments that these differences in performance become negligible when using a folding horizon approach. The aim of this chapter is to convince users of adjustable robust optimization to check for existence of multiple solutions.

In Chapter 6, we introduce a model that treats past data itself as an uncertain model parameter. We show that computational tractability of the robust counterparts associated with this extension of adjustable robust optimization is still maintained. The benefits of the new model are demonstrated by a numerical test case of a well-studied production-inventory problem.

1.4 Disclosure

This thesis is based on the following five research papers:

Chapter 2 D. Bertsimas and F.J.C.T. de Ruiter 2016. Duality in two-stage adaptive linear optimization: faster computation and stronger bounds. *INFORMS Journal on Computing* 28 (3), p500–511. Winner of the INFORMS Optimization Society Student Paper Prize 2017.

Chapter 3 F.J.C.T. de Ruiter, J. Zhen and D. den Hertog 2017. Dual approach for two-stage nonlinear robust optimization. *To be submitted.*

Chapter 4 F.J.C.T. de Ruiter and A. Ben-Tal 2017. Improvement of linear decision rules in robust optimization by lifted uncertainty sets. *To be submitted.*

Chapter 5 F.J.C.T. de Ruiter, R.C.M. Brekelmans and D. den Hertog 2016. The impact of the existence of multiple adjustable robust solutions. *Mathematical Programming* 160 (1), p531–545.

Chapter 6 F.J.C.T. de Ruiter, A. Ben-Tal, R.C.M. Brekelmans and D. den Hertog 2017. Robust optimization of uncertain multistage inventory systems with inexact data in decision rules. *Computational Management Science* 14 (1), p45–77.

Each chapter contains ideas and contributions from all its respective authors. All sections of Chapters 1, 2, 4 and 5 are written by me and experiments are done by me. The sections of Chapter 3 are also written by me except for the text, as well as the numerical experiments, in Sections 3.4, 3.5 and 3.6. Chapter 6 is written by me and corresponding experiments are done by me except for the writing in Section 6.1.

CHAPTER 2

Duality in two-stage adaptive linear optimization: faster computation and stronger bounds

2.1 Introduction

Many applications for decision making under uncertainty can be naturally modeled as two-stage adaptive optimization models. In these models some of the decisions have to be made *here-and-now* before the realization of the uncertain parameter is known. The other decisions are of a *wait-and-see* type, which are chosen after the realization of the uncertain parameter is known. One way of dealing with these problems is via stochastic optimization. These methods assume that a probabilistic description of the realization is known and optimize for expected values. For references on these techniques we refer to Birge and Louveaux (2011) and Kall and Wallace (1994). Stochastic models, especially in a two-stage setting, are known to suffer from the 'curse of dimensionality' and are therefore likely not tractable, see e.g. Shapiro and Nemirovski (2005). A different approach is to model these two-stage problems in a robust setting. Robust optimization techniques do not require a probabilistic description of the uncertainty set and have proven to be very useful in a number of practical applications. A selection of applications that use a two-stage robust setting are: unit commitment in the energy sector (Bertsimas et al. 2013; Wang et al. 2013; Zhao and Zeng 2012), emergency supply chain planning (Ben-Tal et al. 2011b), facility location problems (Ardestani-Jaafari and Delage 2017; Atamtürk and Zhang 2007; Gabrel et al. 2014a), Capacity expansion of network flows (Ordóñez and Zhao 2007; Yin et al. 2009) and many others, see e.g. the survey papers by Bertsimas et al. (2011b) and Gabrel et al. (2014b).

In the last decade or so, there has been a rise in solution techniques tailored to solve two-stage optimization models in a robust setting. One of the first and very popular method is the use of affine policies for the wait-and-see decisions proposed by Ben-Tal et al. (2004). This method is appealing because it is computationally tractable for problem instances of moderate to large size. Furthermore, the affine policies appear to be near optimal in practical applications (Ardestani-Jaafari and

Delage 2017; Ben-Tal et al. 2004; Ben-Tal et al. 2005). The use of affine policies is even provably optimal in some special cases (Bertsimas et al. 2010; Iancu et al. 2013; Gounaris et al. 2013). Other methods designed to solve two-stage adaptive optimization models are: approximation by static solutions (Bertsimas and Goyal 2010), finite adaptability (Bertsimas and Caramanis 2010), enumeration of vertices of the uncertainty set (Bertsimas and Goyal 2012), column generation algorithms (Zeng and Zhao 2013) and iterative partitioning of the uncertainty set (Postek and den Hertog 2016; Bertsimas and Dunning 2016).

In this chapter we derive a new dualized formulation of two-stage adaptive linear models that allow for faster computations and stronger bounds. More specifically, the main contributions of this chapter can be summarized as follows:

1. We provide a dualized two-stage adaptive model for linear two-stage models with continuous wait-and-see decisions. The new model is derived by consecutively dualizing over the wait-and-see decisions and the uncertain parameters. The new dualized formulations have the same set of feasible (and optimal) here-and-now decisions as the original two-stage models. It has different dimensions, uncertain parameters, wait-and-see decisions and constraints than the original two-stage adaptive model. Since the model is again a two-stage adaptive model, all existing solution techniques for two-stage adaptive models can be used to solve it.

2. We show that both formulations also have the same set of feasible and optimal here-and-now decisions when we solve the models using the popular method of affine policies. Furthermore, we show how the original affine policy can be obtained instantly from the affine policy in the dualized formulation.

3. We describe an algorithm to strengthen the lower bound method from Hadjiyiannis et al. (2011) to assess the (sub)optimality of affine policies described using both affine policies from the original and the dualized formulation.

4. We provide empirical evidence that the dualized model in the context of two-stage lot-sizing on a network and two-stage facility location problems solves an order of magnitude faster than the primal formulation with affine policies and provides stronger lower bounds. Furthermore, we provide an explanation and associated empirical evidence that offer insight on which characteristics of the dualized formulation make computations faster.

Our dualized formulation can be used for general two-stage adaptive linear models with both continuous and integer here-and-now decisions. However, since we dualize over the second stage variables, the new dualized formulation only works for continuous second stage decisions. Furthermore, to end up with tractable models, our method focuses on polyhedral uncertainty sets.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the two-stage adaptive optimization model and derive the new dualized two-stage model. We explain the use of affine policies in the primal and dual formulation in Section 2.3. Section 2.4 gives the computational algorithm to obtain stronger bounds on the optimal value of the fully adaptive model. In Sections 2.5 and 2.6, we present our numerical results and show the computational advantage of the dualized formulation. Section 2.7 gives some concluding remarks.

Notation. Throughout this chapter we write vectors and matrices in bold font and scalars in normal font. We use the vector $\boldsymbol{e}$ to denote the vector of all ones and $\boldsymbol{I}$ for the identity matrix. The vector $\boldsymbol{0}$ and matrix $\boldsymbol{O}$ consist of only zero entries. All inequality signs represent componentwise inequalities.

2.2 Duality in two-stage adaptive formulations

We first state the usual two-stage formulation in Section 2.2.1. The new dualized formulation is given in Section 2.2.2. We also indicate similarities in structure with the primal formulation and the differences in the two formulations.

2.2.1 The primal formulation

We consider a general two-stage adaptive optimization model with continuous wait-and-see decisions. In the first stage we set the value of the here-and-now decisions $\boldsymbol{x}$ that have to be decided before the realization of the uncertain parameter is known. The continuous wait-and-see decisions $\boldsymbol{y} \geq \boldsymbol{0}$ have to be chosen after the value of the uncertain parameter is revealed. We take a polyhedral description of the uncertainty set of the form:

$$\mathcal{U} = \{\boldsymbol{\zeta} \geq \boldsymbol{0} : \ \boldsymbol{D}\boldsymbol{\zeta} \leq \boldsymbol{d}\}, \tag{2.1}$$

with $\boldsymbol{D} \in \mathbb{R}^{p \times L}$ and $\boldsymbol{d} \in \mathbb{R}^{p}$. This type of uncertainty sets includes popular sets such as the box-uncertainty and budget uncertainty set (Bertsimas and Sim 2004). The two-stage adaptive optimization problem has a linear objective and a set of linear uncertain constraints. With this general setting we can state the following description of a two-stage linear adaptive optimization model similar to (1.6) introduced in

Chapter 1:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U} : \ \exists \boldsymbol{y} \geq \boldsymbol{0} : \ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.2}$$

where $\mathcal{X} \subset \mathbb{R}^n$ is a set with additional constraints on the here-and-now decisions (some of the $\boldsymbol{x}$ variables may be integer). The wait-and-see variable $\boldsymbol{y}$ has dimension k and we denote the number of constraints in the model by m, so $\boldsymbol{B} \in \mathbb{R}^{m \times k}$. Furthermore, we have $c \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{R} \in \mathbb{R}^{m \times L}$ and $r \in \mathbb{R}^m$. The matrix $\boldsymbol{R}$ is chosen constant in this model, so the model only has uncertainty in the right-hand side. This is mainly done for exposition and all our results can be extended to the case where $\boldsymbol{R}$ depends on the here-and-now decision $\boldsymbol{x}$, for example by taking

$$\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{R}_0 + \sum_{i=1}^{n} \boldsymbol{R}_i x_i,$$

for some matrices $\boldsymbol{R}_0, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_n$. For our dual derivation to work, we must have the matrix $\boldsymbol{B}$ to be fixed independent of $\boldsymbol{\zeta}$. Hence, we only consider the case of *fixed recourse*. Without loss of generality, there is no uncertainty in the objective function and it only includes here-and-now decisions. Objectives including uncertain parameters and wait-and-see decisions can be modeled as an instance of (2.2) using an epigraph formulation, see Ben-Tal et al. (2009, pp. 10-11). These epigraph formulations are also used in the models of our numerical examples in Sections 2.5 and 2.6.

2.2.2 The new dualized formulation

The main contributions of this chapter come from the next theorem, giving a dual formulation of (2.2).

Theorem 2.1 *The here-and-now decision $\boldsymbol{x}$ is feasible (and optimal) for* (2.2) *with nonempty uncertainty set $\mathcal{U}$ as in* (2.1) *if and only if $\boldsymbol{x}$ is feasible (and optimal) for*

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{s.t.} \quad & \forall \boldsymbol{w} \in \mathcal{V} : \ \exists \boldsymbol{\lambda} \geq \boldsymbol{0} : \ \begin{cases} \boldsymbol{w}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{r}) - \boldsymbol{d}^\top \boldsymbol{\lambda} \geq 0 \\ \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w} \end{cases} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.3}$$

where $\mathcal{V} = \left\{ \boldsymbol{w} \geq \boldsymbol{0} : \ \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0}, \ \boldsymbol{e}^\top \boldsymbol{w} = 1 \right\}$.

The proof of this theorem is split in two parts. The first part comes from a result known in the literature and the second part is the new contribution leading to the dualized formulation. The result from the literature transforms (2.2) into a bilinear optimization model by applying duality to the wait-and-see variables. The result from this part is used frequently in the literature, in various settings, to solve two-stage adaptive optimization problems using column generation and Benders decomposition type algorithms (see e.g. Bertsimas et al. (2013), Minoux (2011), Thiele et al. (2009), and Zeng and Zhao (2013) and Zhao and Zeng (2012)) or to derive an exact solution for special cases (Ordóñez and Zhao 2007). This known result is given in Lemma 2.1.

Lemma 2.1 *The here-and-now decision* $\boldsymbol{x}$ *is feasible (and optimal) for* (2.2) *if and only if* $\boldsymbol{x}$ *is feasible (and optimal) for*

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\zeta}\in\mathcal{U}} \max_{\boldsymbol{w}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{w}^\top (\boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) \mid \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0} \right\}. \tag{2.4}$$

Proof. We can write (2.2) as

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\zeta}\in\mathcal{U}} \min_{\boldsymbol{y}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \right\}. \tag{2.5}$$

The result then follows by dualizing over $\boldsymbol{y}$. Note that strong duality for linear programming holds since $\boldsymbol{w} = 0$ is feasible in the resulting model. ■

Note that for every $\boldsymbol{\zeta}$ the variable $\boldsymbol{w}$ ensures that the problem returns ∞ whenever there exists a $\boldsymbol{\zeta}$ that violates the constraints in the original model (2.2). The result from Lemma 2.1 is also used in Kuhn et al. (2011) to assess the suboptimality of affine policies in a two-stage *stochastic* setting. Their bound can also be used in robust settings, but one has to assign a distribution to the uncertainty set a priori. The authors explain that in that case the quality of the bound depends on the a priori distribution that is chosen. For the rest of the proof we first dualize (2.4) further to end up with an equivalent two-stage adaptive optimization formulation.

Proof. *Proof of Theorem 2.1.*

Consider, for fixed $\boldsymbol{w}$, the inner maximization problem in (2.4). Dualizing over $\boldsymbol{\zeta}$ gives

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{w}\geq\boldsymbol{0}} \min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{w}^\top (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top \boldsymbol{\lambda} \mid \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w},\ \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0} \right\} \\ &\qquad = \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{w}\in\widetilde{\mathcal{V}}} \min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{w}^\top (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top \boldsymbol{\lambda} \mid \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w} \right\}, \end{aligned} \tag{2.6}$$

where in the last line we introduced $\widetilde{\mathcal{V}} = \left\{ \boldsymbol{w} \geq \boldsymbol{0} :\ \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0} \right\}$. Note that strong duality for linear programming holds since $\mathcal{U}$ is nonempty. Introducing a variable γ

we write the model using an epigraph formulation

$$\begin{aligned} \min_{\boldsymbol{x},\gamma} \quad & \boldsymbol{c}^\top \boldsymbol{x} + \gamma \\ \text{s.t.} \quad & \forall \boldsymbol{w} \in \widetilde{\mathcal{V}}: \ \exists \boldsymbol{\lambda} \geq \boldsymbol{0}: \ \begin{cases} \boldsymbol{w}^\top(\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top \boldsymbol{\lambda} \leq \gamma \\ \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w} \end{cases} \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned}$$

Now we know that for every feasible solution we must have $\gamma = 0$, since by strong duality the optimal objectives of (2.6) and (2.5) are the same. To end up with our final result (2.3) we have to prove that we can add the additional restriction $\boldsymbol{e}^\top \boldsymbol{w} = 1$ to bound the uncertainty set $\widetilde{\mathcal{V}}$ without affecting the set of feasible solutions. From (2.6) it follows that there has to be an optimal adaptive policy $\boldsymbol{\lambda}^*(\boldsymbol{w})$ that satisfies

$$\boldsymbol{d}^\top \left(\boldsymbol{\lambda}^*(\boldsymbol{w})\right) = \min_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \{\boldsymbol{d}^\top \boldsymbol{\lambda} \mid \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w}\}.$$

Note that $\boldsymbol{d}^\top \left(\boldsymbol{\lambda}^*(\boldsymbol{w})\right)$ is always bounded for fixed w since $\mathcal{U}$ is nonempty. Now, let $t \geq 0$ and $\boldsymbol{w} \geq \boldsymbol{0}$. Then we have

$$\begin{aligned} \boldsymbol{d}^\top \left(\boldsymbol{\lambda}^*(t\boldsymbol{w})\right) &= \min_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \{\boldsymbol{d}^\top \boldsymbol{\lambda} \mid \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top (t\boldsymbol{w})\} \\ &= \min_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \{\boldsymbol{d}^\top (t\boldsymbol{\lambda}) \mid \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w}\} = \boldsymbol{d}^\top \left(t\boldsymbol{\lambda}^*(\boldsymbol{w})\right). \end{aligned}$$

Hence, we can impose scalar multiplicity on the adaptive policy $\boldsymbol{\lambda}^*(\boldsymbol{w})$ without affecting the value of $\boldsymbol{d}^\top \left(\boldsymbol{\lambda}^*(\boldsymbol{w})\right)$. That is, for every $\boldsymbol{w} \in \widetilde{\mathcal{V}}$ and scalar $t \geq 0$ we impose $\boldsymbol{\lambda}^*(t\boldsymbol{w}) = t\boldsymbol{\lambda}^*(\boldsymbol{w})$. Since $\widetilde{\mathcal{V}}$ is a cone, we have that $(t\boldsymbol{w}) \in \widetilde{\mathcal{V}}$ for every $t \geq 0$ and $\boldsymbol{w} \in \widetilde{\mathcal{V}}$. Consider a solution that is feasible for all values in the further restricted uncertainty set

$$\begin{aligned} \mathcal{V} &= \left\{\boldsymbol{w} \geq \boldsymbol{0}: \ \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0}, \ ||\boldsymbol{w}||_1 = 1\right\} \\ &= \left\{\boldsymbol{w} \geq \boldsymbol{0}: \ \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0}, \ \boldsymbol{e}^\top \boldsymbol{w} = 1\right\}. \end{aligned}$$

Then, by scalar multiplicity of $\boldsymbol{\lambda}^*(\boldsymbol{w})$, we can directly construct the other feasible wait-and-see decisions for all other $\boldsymbol{w} \in \widetilde{\mathcal{V}}$ (with $||\boldsymbol{w}||_1 \neq 1$). ■

Any two-stage adaptive optimization model with fixed recourse, continuous wait-and-see decisions and a polyhedral uncertainty set can be readily formulated as an instance of (2.2). Theorem 2.1 then directly provides practitioners with the alternative dual formulation (2.3). Table 2.1 highlights some differences such as the number of wait-and-see variables, uncertain parameters and constraints in the primal and dual formulation. In our numerical examples in Sections 2.5 and 2.6 we clarify these differences with explicit values for m, k, L and p.

Table 2.1 – Comparing dimensions of uncertainty parameters, variables and number of constraints in the original two-stage adaptive formulation (2.2) and in our new dualized formulation (2.3).

	Primal formulation (2.2)	Dual formulation (2.3)
# uncertain parameters	L	m
# wait-and-see decisions	k	p
# constraints on variables	m	$L+1$
# constraints on uncertain parameter	p	$k+1$

2.3 Solving the primal and dual formulation with affine policies

The model (2.3) is again a two-stage adaptive robust optimization model with a nonnegative bounded polyhedral uncertainty set and is therefore another instance of (2.2). Hence, we can directly apply all exact and approximation methods to solve adaptive optimization problems mentioned in the introduction. We first show the equivalence of the dual formulation with the nonadaptive robust counterpart in the static case. We then continue to show that the optimal solutions of both formulations are the same when we solve the models with affine policies.

2.3.1 Static robust optimization

If we take $\boldsymbol{B} = \boldsymbol{O}$, then (2.2) is the following robust optimization model without wait-and-see decisions:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U}: \; \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.7}$$

where $\mathcal{U}$ is as in (2.1). This problem is hard to solve in its current form since each constraint has to hold for an infinite number of values for $\boldsymbol{\zeta}$. To reformulate the problem, we can consider the uncertainty constraintwise (see Ben-Tal et al. (2009)), i.e., we only have to look at one row

$$\forall \boldsymbol{\zeta} \in \mathcal{U}: \; \boldsymbol{A}_i \boldsymbol{x} \geq \boldsymbol{R}_i \boldsymbol{\zeta} + r_i \tag{2.8}$$

at a time, where $\boldsymbol{A}_i, \boldsymbol{R}_i$ and r_i are respectively the i-th row of $\boldsymbol{A}, \boldsymbol{R}$ and $\boldsymbol{r}$. To make this model tractable we can reformulate each constraint using standard duality techniques to obtain the robust counterpart, see e.g. Ben-Tal et al. (2009).

Lemma 2.2 (Robust Counterpart) *Constraint* (2.8) *is satisfied if and only if* *there exists a* $\boldsymbol{\pi^i} \in \mathbb{R}^p$ *such that*

$$\begin{aligned}
&\boldsymbol{A}_i\boldsymbol{x} - \boldsymbol{\pi^i}^\top \boldsymbol{d} \geq r_i \\
&\boldsymbol{D}^\top \boldsymbol{\pi^i} \geq \boldsymbol{R}_i \\
&\boldsymbol{\pi^i} \geq \boldsymbol{0}.
\end{aligned}$$

Note that this dualization approach can also be used for any other polyhedral uncertainty set. For notational convenience we shall use matrix variables for the rest of the section. If we write $\boldsymbol{\Pi} = [\boldsymbol{\pi^i}, \ldots, \boldsymbol{\pi^m}]$, then by Lemma 2.2 we have that (2.7) is equivalent to

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{\Pi}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{\Pi}^\top \boldsymbol{d} \geq \boldsymbol{r} \\
& \boldsymbol{D}^\top \boldsymbol{\Pi} \geq \boldsymbol{R}^\top \\
& \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\Pi} \geq \boldsymbol{O}.
\end{aligned} \tag{2.9}$$

We can also find a dual formulation for the static model (2.7) using the dual formulation that is derived in Theorem 2.1. In that way, we end up with the same dual formulation as in (2.3), but with the simple uncertainty set

$$\mathcal{V} = \left\{ \boldsymbol{w} \geq \boldsymbol{0} : \ \boldsymbol{e}^\top \boldsymbol{w} = 1 \right\}. \tag{2.10}$$

For these robust models with $\boldsymbol{B} = \boldsymbol{O}$ the uncertainty set (2.10) has only m extreme points $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_m}$. As shown in Bertsimas and Goyal (2012, Lemma 1), linear policies are optimal if there are only m extreme points in the uncertainty set, where m is the dimension of the uncertain parameter. Furthermore, by taking the linear policy $\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{\Pi}\boldsymbol{w}$ in (2.3) we end up with the same robust counterpart as (2.9).

2.3.2 Solving the two-stage formulations with affine policies

Let us now return to the general case in which $\boldsymbol{B} \neq \boldsymbol{O}$, so we do need to take the wait-and-see decisions $\boldsymbol{y}$ into account. In principle, an optimal policy $\boldsymbol{y}(\boldsymbol{\zeta})$ in (2.2) can be any function of the uncertain parameter $\boldsymbol{\zeta}$. However, this results in an intractable model where we would have to optimize over all possible functions. To come up with tractable models Ben-Tal et al. (2004) suggest to restrict the wait-and-see decisions to be affine in $\boldsymbol{\zeta}$:

$$\boldsymbol{y}(\boldsymbol{\zeta}) = \boldsymbol{u} + \boldsymbol{V}\boldsymbol{\zeta},$$

where $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{V} \in \mathbb{R}^{m \times L}$ are respectively a vector and a matrix of here-and-now variables. Although this restriction might seem very severe, it turns out to perform very good in practical applications, see Ben-Tal et al. (2004) and Ben-Tal et al. (2005), and is even provably optimal in some specific cases, see Bertsimas et al. (2010) and Iancu et al. (2013). With this decision rule, we obtain the following robust counterpart for (2.2) with affine policies

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{V}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U} : \begin{cases} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{u} + \boldsymbol{V}\boldsymbol{\zeta}) \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \\ \boldsymbol{u} + \boldsymbol{V}\boldsymbol{\zeta} \geq \boldsymbol{0} \end{cases} \\
& \boldsymbol{x} \in \mathcal{X},
\end{aligned} \tag{2.11}$$

which is similar to model (1.8) stated in Chapter 1. This model does not have wait-and-see variables. Therefore, we can apply Lemma 2.2 to reformulate each constraint and obtain the robust counterpart. Introducing the auxiliary (matrix) variables $\boldsymbol{\Pi} \in \mathbb{R}^{p\times m}$ and $\boldsymbol{\Xi} \in \mathbb{R}^{p\times k}$ we can write down the robust counterpart as

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{V},\boldsymbol{\Pi},\boldsymbol{\Xi}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} - \boldsymbol{\Pi}^\top \boldsymbol{d} \geq \boldsymbol{r} \\
& \boldsymbol{B}\boldsymbol{V} \geq \boldsymbol{R} - \boldsymbol{\Pi}^\top \boldsymbol{D} \\
& \boldsymbol{u} - \boldsymbol{\Xi}^\top \boldsymbol{d} \geq \boldsymbol{0} \\
& \boldsymbol{D}^\top \boldsymbol{\Xi} + \boldsymbol{V}^\top \geq \boldsymbol{O} \\
& \boldsymbol{\Pi}, \boldsymbol{\Xi} \geq \boldsymbol{O} \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned} \tag{2.12}$$

For the dualized formulation we can also impose linear restrictions, i.e.,

$$\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{Q}\boldsymbol{w}, \tag{2.13}$$

where we now introduce here-and-now variables $\boldsymbol{Q} \in \mathbb{R}^{p\times m}$ to construct the decision rule. Note that we restricted ourselves now to *linear* policies in the dual formulation instead of affine policies. However, leaving out the constant term does not restrict the set of feasible and optimal here-and-now decisions as follows from the next proposition.

Proposition 2.1 *If* $(\boldsymbol{x}, \boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{q} + \boldsymbol{Q}\boldsymbol{w})$ *is feasible for* (2.3), *then* $(\boldsymbol{x}, \widetilde{\boldsymbol{\lambda}}(\boldsymbol{w}) = \widetilde{\boldsymbol{Q}}\boldsymbol{w})$ *with* $\widetilde{\boldsymbol{Q}} = \boldsymbol{q}\boldsymbol{e}^\top + \boldsymbol{Q}$ *is also feasible.*

Proof. For all $\boldsymbol{w} \in \mathcal{V}$ we have $\boldsymbol{e}^\top \boldsymbol{w} = 1$. Therefore, for all $\boldsymbol{w} \in \mathcal{V}$ the following relation holds

$$\widetilde{\boldsymbol{\lambda}}(\boldsymbol{w}) = \widetilde{\boldsymbol{Q}}\boldsymbol{w} = (\boldsymbol{q}\boldsymbol{e}^\top + \boldsymbol{Q})\boldsymbol{w} = \boldsymbol{q} + \boldsymbol{Q}\boldsymbol{w} = \boldsymbol{\lambda}(\boldsymbol{w}).$$

Hence, if $\boldsymbol{\lambda}(\boldsymbol{w})$ is a feasible policy for (2.14), then so is $\widetilde{\boldsymbol{\lambda}}(\boldsymbol{w})$. ■

Substituting the linear policy (2.13) in (2.3), we obtain the following model

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{Q}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \forall \boldsymbol{w} \in \mathcal{V} : \begin{cases} \boldsymbol{w}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{r}) - \boldsymbol{d}^\top(\boldsymbol{Q}\boldsymbol{w}) \geq \boldsymbol{0} \\ \boldsymbol{D}^\top \boldsymbol{Q}\boldsymbol{w} \geq \boldsymbol{R}^\top \boldsymbol{w} \\ \boldsymbol{Q}\boldsymbol{w} \geq \boldsymbol{0} \end{cases} \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned} \tag{2.14}$$

A robust counterpart for (2.14) can be derived using standard LP dualization as in Lemma 2.2. With the introduction of the auxiliary variables $\boldsymbol{\varepsilon} \in \mathbb{R}^k$, $\boldsymbol{\Lambda} \in \mathbb{R}^{k \times L}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{k \times p}$, the resulting robust counterpart can be written as

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{Q},\boldsymbol{\varepsilon},\boldsymbol{\Omega},\boldsymbol{\Lambda}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{\varepsilon} - \boldsymbol{Q}^\top \boldsymbol{d} \geq \boldsymbol{r} \\
& \boldsymbol{B}\boldsymbol{\Lambda} \geq \boldsymbol{R} - \boldsymbol{Q}^\top \boldsymbol{D} \\
& \boldsymbol{B}\boldsymbol{\Omega} + \boldsymbol{Q}^\top \geq \boldsymbol{O} \\
& \boldsymbol{\varepsilon} \geq \boldsymbol{0}, \ \boldsymbol{\Lambda}, \boldsymbol{\Omega} \geq \boldsymbol{O} \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned} \tag{2.15}$$

The next theorem shows that the primal and dual formulation have the same set of feasible (and optimal) here-and-now decisions.

Theorem 2.2 *The solution* $(\boldsymbol{x}, \boldsymbol{Q}, \boldsymbol{\varepsilon}, \boldsymbol{\Omega}, \boldsymbol{\Lambda})$ *is feasible for* (2.15) *if and only if the solution* $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{V}, \boldsymbol{\Pi}, \boldsymbol{\Xi})$ *is feasible for* (2.12)*, where*

$$\begin{aligned}
\boldsymbol{u} &= \boldsymbol{\varepsilon} + \boldsymbol{\Omega}\boldsymbol{d} \\
\boldsymbol{V} &= \boldsymbol{\Lambda} - \boldsymbol{\Omega}\boldsymbol{D} \\
\boldsymbol{\Pi} &= \boldsymbol{\Omega}^\top \boldsymbol{B}^\top + \boldsymbol{Q} \\
\boldsymbol{\Xi} &= \boldsymbol{\Omega}^\top.
\end{aligned}$$

The proof is direct and therefore omitted. Theorem 2.2 is not only useful because it proves equivalence of the primal and dual formulation with affine policies. It also allows us to solve the dual formulation (2.15) with affine policies and directly obtain the optimal affine policy of the original formulation (or vice versa). Despite this equivalence there may be significant computational benefits from solving two stage problems using the dualized formulation rather than the primal formulation. This can be seen by comparing the two robust counterparts (2.12) and (2.15). We compare the number of affine constraints and the number of sign restrictions in Table 2.2. We use the same parameters as in Table 2.1 for the number of uncertain parameters (L), the number of wait-and-see decisions (k), the number of affine constraints on the variables (m) and the number of affine constraints in the uncertainty set (p). We

Table 2.2 – Comparing the number of affine constraints and sign restrictions in (2.12) and (2.15)

	Primal formulation (2.12)	Dual formulation (2.15)
# affine constraints	$(1 + L)(m + k)$	$m(1 + L + p)$
# sign restrictions	$p(m + k)$	$k(1 + L + p)$

observe that the total number of constraints (affine constraints and sign restrictions) is the same in both formulations. However, there is a difference in the break down into the number of affine constraints and the number of sign restrictions. This is important since sign restrictions are much easier to handle by solvers than affine constraints. From Table 2.2 we see that for a large number of wait-and-see decisions k, relative to the number of constraints in the original model and uncertainty set (m and p), the dual formulation (2.15) can most likely be solved more efficiently than the primal formulation (2.12). We observe these computational benefits in our numerical examples in Sections 2.5 and 2.6 where we present Table 2.2 with some explicit values for L, k, m and p.

Finally, we note that the models (2.11) and (2.14) can also be solved via cutting plane methods, see Mutapcic and Boyd (2009). There have been extensive numerical studies that show that in some cases cutting plane algorithms require slightly less computation time than solving the robust counterpart constructed by Lemma 2.2 (Fischetti and Monaci 2012; Bertsimas et al. 2016). We have also solved our numerical examples with the cutting plane algorithm described in those papers. As with the reformulation approach, we observe that the dual formulation (2.14) can be solved an order of magnitude faster than the primal problem. This approach is however not elaborated further for two reasons. First, to construct the primal solution from the dual solution by Theorem 2.2 we need the auxiliary variables that are introduced by the reformulation. Second, initial findings showed that the cutting plane algorithm is a lot slower for the problems considered Sections 2.5 and 2.6. We were only able to solve the smaller instances in reasonable time via cutting planes.

2.4 Stronger bounds on the optimality gap of affine policies

In general, the restriction from fully adaptive policies to affine policies is both for the primal and dual formulation an approximation of the fully adaptive solution. It is important to provide methods that can efficiently determine bounds on the (sub)optimality of affine policies. Here we extend a method that was first presented in Hadjiyiannis et al. (2011) to provide bounds on the optimality gap of affine policies. We first explain the initial idea from Hadjiyiannis et al. (2011) and then describe the algorithm that provides stronger bounds.

The main idea is to solve the fully adjustable model (2.2) only for a finite subset of the uncertainty set. Clearly, any optimal solution to this model results in a lower bound since we only guarantee feasibility for a strict subset of the uncertainty region. If we denote the finite subset by $\left\{\boldsymbol{\zeta^1}, \boldsymbol{\zeta^2}, \ldots, \boldsymbol{\zeta^{\overline{N}}}\right\}$, then we end up with the following equivalent deterministic optimization model

$$
\begin{aligned}
\min_{\boldsymbol{x}, \boldsymbol{y^1}, \ldots, \boldsymbol{y^{\overline{N}}}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y^i} \geq \boldsymbol{R}\boldsymbol{\zeta^i} + \boldsymbol{r} \qquad \forall i = 1, \ldots, \overline{N} \\
& \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y^1}, \ldots, \boldsymbol{y^{\overline{N}}} \geq 0.
\end{aligned}
\tag{2.16}
$$

The crucial question is of course which scenarios to include. It is shown by Bertsimas and Goyal (2012) that the lower bound is tight if we include all extreme points of the uncertainty set. This is in practice undoable since there can be a huge number of extreme points, each resulting in an extra variable and constraint in (2.16). Another straightforward way would be to sample $\overline{N}$ scenarios uniformly at random from $\mathcal{V}$. The model (2.16) remains tractable for relatively large $\overline{N}$, but for all our examples we obtain useless bounds, even when the number of random samples $\overline{N}$ is as big as 10^5. We therefore have to pick the scenarios in a more specific way. To do so, we first introduce the notion of *binding scenarios.*

Definition 2.1 (Binding scenarios) *Let $f : \mathcal{U} \times \mathcal{X} \to \mathbb{R}$ be a function of the uncertain parameter $\boldsymbol{\zeta} \in \mathcal{U}$ and here-and-now decision $\boldsymbol{x} \in \mathcal{X}$. For a given $\boldsymbol{x} \in \mathcal{X}$ the parameter $\widehat{\boldsymbol{\zeta}}$ is called* binding *for the robust constraint*

$$
f(\boldsymbol{\zeta}, \boldsymbol{x}) \leq 0 \qquad \forall \zeta \in \mathcal{U}
$$

if $f(\widehat{\boldsymbol{\zeta}}, \boldsymbol{x}) = 0$.

In the primal formulation with affine policies we only have here-and-now decisions $\boldsymbol{x}, \boldsymbol{u}$ and $\boldsymbol{V}$. Furthermore, each robust constraint is linear in the here-and-now decision and the uncertain parameter. Therefore, a binding scenario can easily be found for each constraint by solving a small linear optimization model $\widehat{\boldsymbol{\zeta}} = \arg\max_{\boldsymbol{\zeta}\in\mathcal{U}} f(\boldsymbol{\zeta}, \boldsymbol{x})$ and check whether the maximum is equal to zero (up to a certain precision). The hope is that scenarios that are binding the solution with affine policies are also binding the fully adaptive solution.

The method by Hadjiyiannis et al. (2011) only uses the information derived from the primal formulation with affine policies (2.2). Using Theorem 2.2 we can directly construct the optimal affine policy in the dual formulation once the optimal affine policy in the primal formulation is known. Using this other affine policy we can construct another subset of $\mathcal{V}$ consisting of binding scenarios in the dual formulation. The resulting deterministic model of the dual formulation with a finite subset $\{\boldsymbol{w^1}, \boldsymbol{w^2}, \ldots, \boldsymbol{w^{\overline{M}}}\}$ is given by

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\lambda^1},\ldots,\boldsymbol{\lambda^{\overline{M}}}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & (\boldsymbol{w^j})^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{r}) - \boldsymbol{d}^\top \boldsymbol{\lambda^j} \geq \boldsymbol{0} \qquad \forall j = 1, \ldots, \overline{M} \\ & \boldsymbol{D}^\top \boldsymbol{\lambda^j} \geq \boldsymbol{R}^\top \boldsymbol{w^j} \qquad \forall j = 1, \ldots, \overline{M} \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\lambda^1}, \ldots, \boldsymbol{\lambda^{\overline{M}}} \geq \boldsymbol{0}. \end{aligned} \tag{2.17}$$

Combining the constraints from (2.16) and (2.17) results in a model that provides a stronger lower bound than the one that only uses the binding scenarios from the primal formulation. We can now give Algorithm 1 that provides the strengthened bound on the optimal value of the fully adaptive model. Step 1 provides us with a feasible solution and an upper bound on the optimal value of the fully adaptive problem. The objective value of the model in step 4 gives us the new lower bound. A binding scenario for each constraint in (2.11) and (2.14) can be found directly using the optimal affine policies from step 1 and 2. We omit here the elaborate description of a more efficient way to finding the set of binding scenarios in step 3 via KKT conditions which is described in Hadjiyiannis et al. (2011). However, step 3 is not the most time consuming step as solving the model with affine policies in step 1 takes by far the most time. Finally, we note that we can also solve the dual formulation (2.15) with affine policies in step 1 and obtain the primal affine policy in step 2 using Theorem 2.2.

2.5 Example 1: lot-sizing on a network

In this section we present a natural example in which (2.14) takes an order of magnitude less time to solve than the primal formulation (2.11). Also, the new lower bound on the fully adaptive model (2.2) derived from Algorithm 1 is much stronger than the lower bound from Hadjiyiannis et al. (2011) that only used the binding scenarios from the primal formulation.

Algorithm 1 Stronger bounds on optimality of affine policies

1: Solve (2.12) to get optimal here-and-now $\boldsymbol{x}$, affine policy $\boldsymbol{y}(\boldsymbol{\zeta}) = \boldsymbol{u} + \boldsymbol{V}\boldsymbol{\zeta}$ and auxiliary variables $\boldsymbol{\Pi}, \boldsymbol{\Xi}$.
2: Construct the dual affine policy $\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{Q}\boldsymbol{w}$ using Theorem 2.2.
3: Find the binding scenarios $\{\boldsymbol{\zeta^1}, \boldsymbol{\zeta^2}, \ldots, \boldsymbol{\zeta^{\overline{N}}}\}$ in (2.11) and $\{\boldsymbol{w^1}, \boldsymbol{w^2}, \ldots, \boldsymbol{w^{\overline{M}}}\}$ in (2.14).
4: Solve the sampled problem with binding scenarios for the primal and dual

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{y^1},\ldots,\boldsymbol{y^{\overline{N}}},\boldsymbol{\lambda^1},\ldots,\boldsymbol{\lambda^{\overline{M}}}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y^i} \geq \boldsymbol{R}\boldsymbol{\zeta^i} + \boldsymbol{r} && \forall i = 1, \ldots, \overline{N} \\
& \boldsymbol{y^1}, \ldots, \boldsymbol{y^{\overline{N}}} \geq 0 \\
& (\boldsymbol{w^j})^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{r}) - \boldsymbol{d}^\top \boldsymbol{\lambda^j} \geq \boldsymbol{0} && \forall j = 1, \ldots, \overline{M} \\
& \boldsymbol{D}^\top \boldsymbol{\lambda^j} \geq \boldsymbol{R}^\top \boldsymbol{w^j} && \forall j = 1, \ldots, \overline{M} \\
& \boldsymbol{\lambda^1}, \ldots, \boldsymbol{\lambda^{\overline{M}}} \geq \boldsymbol{0} \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned}$$

2.5.1 Problem setting

In lot-sizing on a network we have to determine the stock allocation x_i for $i = 1, \ldots, N$ stores prior to knowing the realization of the demand at each location. The demand $\boldsymbol{\zeta}$ is uncertain and assumed to be in a budget uncertainty set:

$$\mathcal{U} = \left\{ \boldsymbol{\zeta} : \ \boldsymbol{0} \leq \boldsymbol{\zeta} \leq \hat{\zeta}\boldsymbol{e}, \ \boldsymbol{e}^\top \boldsymbol{\zeta} \leq \Gamma \right\}. \tag{2.18}$$

After we observe the realization of the demand we can transport stock y_{ij} from store i to store j at cost t_{ij} in order to meet all demand. The aim is to minimize the worst case storage costs (with unit costs c_i) and the cost arising from shifting the products from one store to another. This network flow model can now be written as a specific instance of the primal problem (2.2) as follows:

$$\begin{aligned}
\min_{\boldsymbol{x},\alpha} \quad & \alpha \\
\text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U} : \exists \boldsymbol{y} \geq \boldsymbol{0} : \left\{ \begin{array}{ll} \alpha \geq \sum_{i=1}^N c_i x_i + \sum_{i=1}^N \sum_{j=1}^N t_{ij} y_{ij} & \\ \zeta_i \leq \sum_{j=1}^N y_{ji} - \sum_{j=1}^N y_{ij} + x_i & i = 1, \ldots, N \end{array} \right. \\
& \qquad\qquad 0 \leq x_i \leq K_i \qquad i = 1, \ldots, N,
\end{aligned} \tag{2.19}$$

where the first line in (2.19) is for the epigraph formulation. The second line contains the balance equations: we have to shift stock to and from node i such that the initial storage plus the net shift in stock still exceeds the demand at node i. The last constraints restrict the capacity of the stock at each node. Note that this model can be seen as a network flow model with multiple sources and multiple sinks.

2.5.2 Test case and numerical results

We pick $N \in \{10, 20, 30, \dots, 100\}$ locations uniformly at random from $[0, 10]^2$. Let t_{ij}, the cost to transport one unit of demand from location i to j, be the Euclidean distance and the unit storage cost c_i be equal to 20. The individual maximum demand $\hat{\zeta}$ and the capacity K_i of each store is set to 20 units. The total demand in the network is set to $\Gamma = 20\sqrt{N}$. This is to avoid trivial and unrealistic cases where either all demand can occur at a single store ($\Gamma = 20$) or where the demand in each store is independent ($\Gamma = 20N$). All computations were carried out with Gurobi 6.0.3 (Gurobi Optimization 2015) on an Intel i7-4770 3.40GHz Windows computer with 8GB of RAM. All modeling was done using the modeling language JuMP (Lubin and Dunning 2015).

We solve both (2.12) and (2.14) and depict the average solution times over 10 runs in Table 2.3, as well as the objective value and the lower bounds. The stock allocation (the here-and-now decision) for the $N = 30$ instance is depicted in Figure 2.1. The

Table 2.3 – Compare performance of primal and dualized formulation with affine policies for the lot-sizing example. The percentages in the last columns depict the optimality gap derived from each lower bound compared to the objective value. All results are averaged over 10 runs.

	Solver time (sec)		Objective value	Lower Bound (Gap%)	
N	Primal	Dual		Primal	Primal/Dual
10	< 0.1	< 0.1	928	797 (**14.0%**)	824 (**11.1%**)
20	0.3	0.1	1353	1113 (**17.7%**)	1190 (**12.0%**)
30	2.6	0.8	1670	1356 (**18.8%**)	1465 (**12.3%**)
40	11.8	2.6	1947	1562 (**19.8%**)	1728 (**11.3%**)
50	42.0	7.3	2188	1728 (**21.0%**)	1934 (**11.6%**)
60	142.2	20.5	2421	1912 (**21.0%**)	2160 (**10.8%**)
70	366.0	41.3	2598	1996 (**23.2%**)	2312 (**11.0%**)
80	826.9	88.7	2781	2136 (**23.2%**)	2495 (**10.3%**)
90	1647.1	179.8	2953	2252 (**23.8%**)	2641 (**10.6%**)
100	4026.2	231.0	3130	2408 (**23.1%**)	2799 (**10.6%**)

lower bound from the primal is obtained using the method from Hadjiyiannis et al. (2011). The primal/dual bound is the strengthened bound resulting from Algorithm

1. Solving the model via the new dualized formulation (2.15) reduces the computation an order of magnitude compared with the original primal formulation (2.12). For the larger instances we see that the primal formulation is approximately 20 times slower. These results are averaged over 10 runs to avoid random peak performances, but in each individual run we observed the significant decrease in computation time. The strengthened primal/dual bound from Algorithm 1 is much tighter than the primal bound, more than halving the optimality gap for the larger instances.

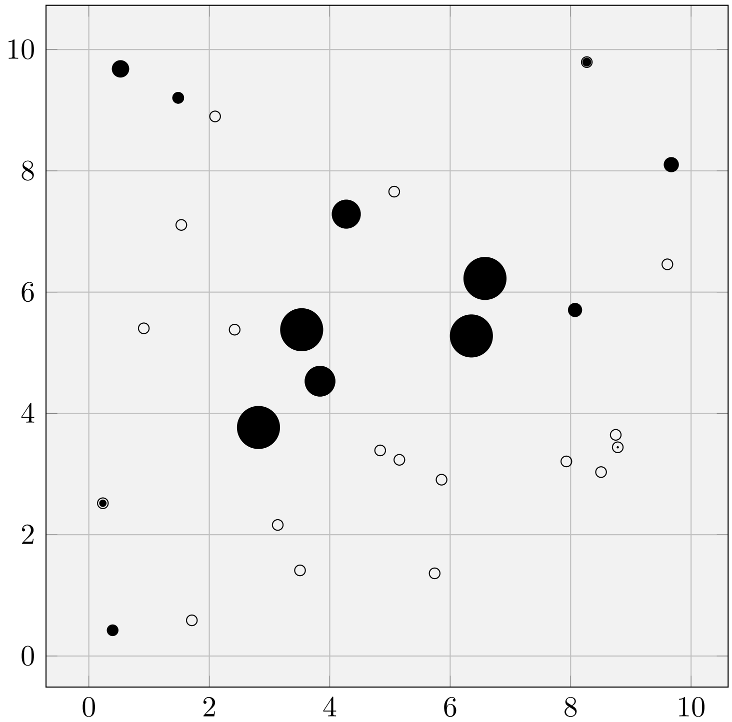


Figure 2.1 – Stock allocation for an instance with 30 stores on the grid $[0,10]^2$. The filled dots have stock and the larger the dots are, the more stock is allocated. The open dots are stores that do not have any stock allocated.

2.5.3 Why is the dual formulation faster?

To understand the significant faster computation time of the dual formulation displayed in Table 2.3, we look at the dimensions (number of uncertain parameters, wait-and-see decisions, constraints on variables and constraints on uncertain parameters) for the case $N = 20$. We give the values of these dimensions in Table 2.3 using the same format as is in Table 2.1. We observe that the primal and dual formulation have the same characteristics, except for the number of wait-and-see decisions and the number of constraints on the uncertain parameter in the uncertainty set. Given these values, we can explicitly calculate the number of affine constraints and the

Table 2.4 – Comparing dimensions of variables, uncertainties and number of constraints in the primal and dual formulation for the lot-sizing instance with $N = 20$ stores.

	Primal formulation (2.2)	Dual formulation (2.3)
# uncertain parameters	20	21
# wait-and-see decisions	400	21
# constraints on variables	21	21
# constraints on uncertain parameter	21	401

number of sign restrictions using the formulas from Table 2.2. The resulting number of constraints and sign restrictions are given in Table 2.5. We observe that the

Table 2.5 – Comparing the number of affine constraints and sign restrictions in (2.12) and (2.15) for the lot-Sizing instance with $N = 20$ stores.

	Primal formulation (2.12)	Dual formulation (2.15)
# affine constraints	8841	882
# sign restrictions	8841	16800

primal formulation (2.12) has about 50 times more affine constraints than the dual formulation (2.15). The dual formulation does have a lot more sign restrictions on its variables, but these are significantly simpler for solvers. To investigate the claim that the number of affine constraints is indeed the cause of the speedup we adapt the $N = 20$ instance from the network lot-sizing model (2.19). From Table 2.2 we see that increasing p, the number of affine constraints in the uncertainty set $\mathcal{U}$, leads to an increase of affine constraints in the dual formulation with affine policies. At the same time, the value of p does not affect the number of affine constraints in the primal formulation. To increase p, we add nonreduntant constraints of the following type to the polyhedral description of $\mathcal{U}$:

$$\sum_{i \in S} \zeta_i \leq 20\sqrt{|S|},$$

where $S \subset \{1, \ldots, N\}$ is a random subset of size $\frac{1}{2}N$. The number of constraints p can be increased at will by adding more of these constraints. Note that increasing p also increases the total number of variables and the number of sign constraints, but these grow in more or less the same order of magnitude in both formulations. If we consider the case $N = 20$, then we find that the number of affine constraints in (2.12) and (2.15) is equal when the number of constraints in the uncertainty set $\mathcal{U}$ equals $p = 400$. Note that $p \geq 21$, since we need 21 constraints to describe the budget uncertainty set. The case with $p = 21$ is therefore just our original network lot-sizing problem (2.19). We measure the difference in computation time between the primal and the dual formulation by the quotient

$$\frac{\textit{Solver time for } (2.12)}{\textit{Solver time for } (2.15)}.$$

In Figure 2.2, we plotted this quotient for $p \in \{21, 22, \ldots, 1000\}$ constraints in the uncertainty set for each random instance. We already know from Table 2.3 that the

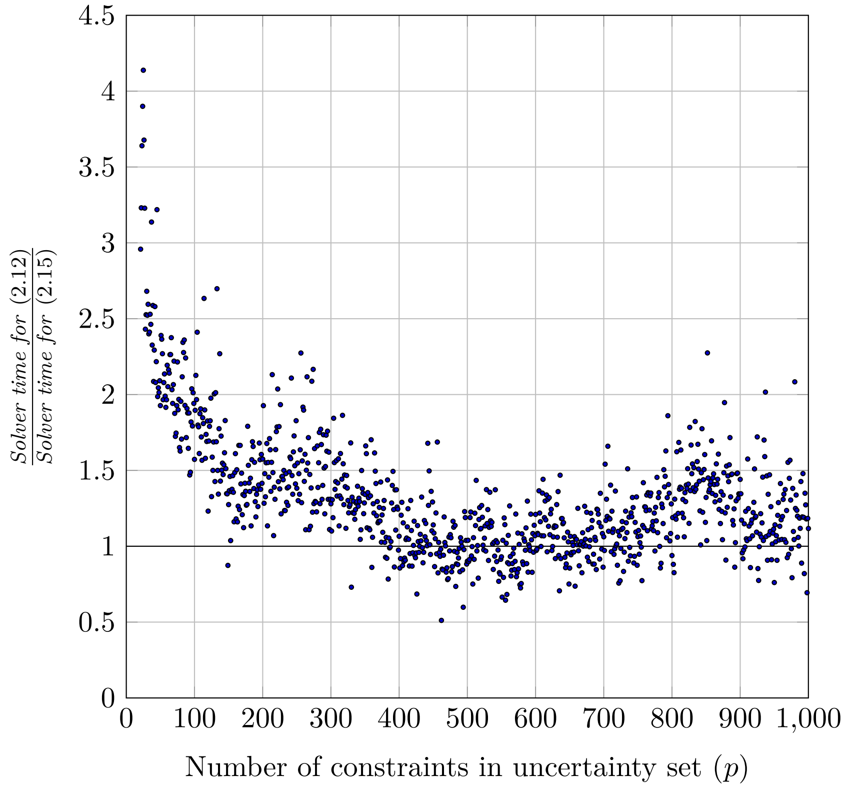


Figure 2.2 – The computation time of the primal formulation (2.12) divided by the time needed to solve the primal formulation (2.12) for $N = 20$ and various number of affine constraints in the uncertainty set p. For $p = 400$ both (2.12) and (2.15) have the same number of affine constraints. Values above the horizontal line at 1 indicate that the dual formulation is solved faster than the primal formulation and vice versa for values smaller than 1.

dual formulation with affine policies solves the original instance three or more times faster than the primal formulation. If we start adding constraints, the computational advantage progressively decreases and after a point it disappears.

2.6 Example 2: facility location problem

The second example we consider is a facility location problem that has also been studied in Ardestani-Jaafari and Delage (2017) and Baron et al. (2011). Similar two-stage adaptive models can be found in Zeng and Zhao (2013). In our results we again observe a significant reduction in computational time required for solving the dualized formulation with affine policies over the primal formulation with affine policies. For this problem, however, the strengthened bounds from Algorithm 1 only slightly improve the bounds obtained from the primal formulation.

2.6.1 Problem setting

We consider a facility location problem where we can build factories at candidate sites $s \in \mathcal{S} = \{1, \dots, S\}$, which have to serve customers $c \in \mathcal{C} = \{1, \dots, C\}$ in the area. The uncertain demand for customer c is modeled as $(1 - \zeta_c)\bar{d}_c$, with $\bar{d}_c$ the nominal demand of customer c and ζ_c the uncertain shock in the demand. We take again a budget uncertainty set of the form

$$\mathcal{U} = \left\{ \boldsymbol{\zeta} : \ \mathbf{0} \leq \boldsymbol{\zeta} \leq \hat{\zeta}\boldsymbol{e}, \ \boldsymbol{e}^\top \boldsymbol{\zeta} \leq \Gamma \right\},$$

where Γ is our budget parameter[1]. There are two types of decisions in this model. First, strategic here-and-now decisions that have to be decided before the demand is known. We have a binary variable x_s to decide whether the facility at site s is opened and a continuous variable p_s to set the capacity level at each opened facility site. Second, wait-and-see decisions y_{sc} on the production at facility s which is transported to customer c. Each unit of demand can generate a revenue of η. There are also several costs incurred for the various strategic and operations decisions. Opening a facility s has a fixed cost f_s and a cost of b_s per unit of capacity installed. The production of one unit at facility s has cost g_s and transporting the goods to customer c bears an additional cost h_{sc}. The goal is to maximize the total profit. This problem can be modeled as a two-stage adaptive optimization model, see Ardestani-Jaafari and Delage (2017):

$$\begin{aligned} \max_{t,\boldsymbol{x},\boldsymbol{p}} \quad & \alpha - \sum_{s \in \mathcal{S}} (b_s p_s + f_s x_s) \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U} : \quad \exists \boldsymbol{y} \geq 0 : \begin{cases} \sum_{s \in \mathcal{S}, c \in \mathcal{C}} (\eta - g_s - h_{sc})\, y_{sc} \geq \alpha \\ \sum_{c \in \mathcal{C}} y_{sc} \leq p_s \qquad \forall s \in \mathcal{S} \\ \sum_{s \in \mathcal{S}} y_{sc} \leq \bar{d}_c - \zeta_c \bar{d}_c \qquad \forall c \in \mathcal{C} \end{cases} \\ & \boldsymbol{p} \leq M\boldsymbol{x}, \ \boldsymbol{x} \in \{0,1\}^N. \end{aligned} \tag{2.20}$$

Note that we have a maximization objective, but this can easily be turned into a minimization objective by the relation $\max_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) = -\min_{\boldsymbol{x} \in \mathcal{X}} (-f(\boldsymbol{x}))$ before applying Theorem 2.1.

[1] In fact, Ardestani-Jaafari and Delage (2017) also consider negative values of the uncertainty parameter. It is not hard to see that these are nonbinding scenarios and we can therefore use this uncertainty set instead.

2.6.2 Test case and numerical results

We consider the same setting as in Ardestani-Jaafari and Delage (2017), which is based on the set-up of an earlier paper on robust facility location planning by Baron et al. (2011). We randomly generate C customers and S sites on a unit square. For the cost parameters we take $f_s = 50000$, $b_s = 0.1$, $g_s = 0.1$, $\eta = 1$. The nominal demand is drawn uniformly at random from $[17500, 22500]$ and $\hat{\zeta} = 0.15$. The transportation cost t_{ij} is just the Euclidean distance between two points i and j. We take $S = 10$ possible sites and $C \in \{10, 20, 30, 40, 50\}$. The cases with $C = 10$ and $C = 20$ are in Ardestani-Jaafari and Delage (2017) referred to as small and medium instances. For the larger instances the computational time vastly increased and they did not report results on the models with affine policies. We use the same computer and optimization software as mentioned in Section 2.5.

The results for various numbers of customers C and various percentage levels of uncertainty Γ are given in Table 2.6. We use the standard notion of budget uncertainty where a budget of 30% means that 30% of the uncertain prameters can be at their extreme value of $\hat{\zeta} = 0.15$. A graph indicating the location and the facilities that are opened for one case is given in Figure 2.3.

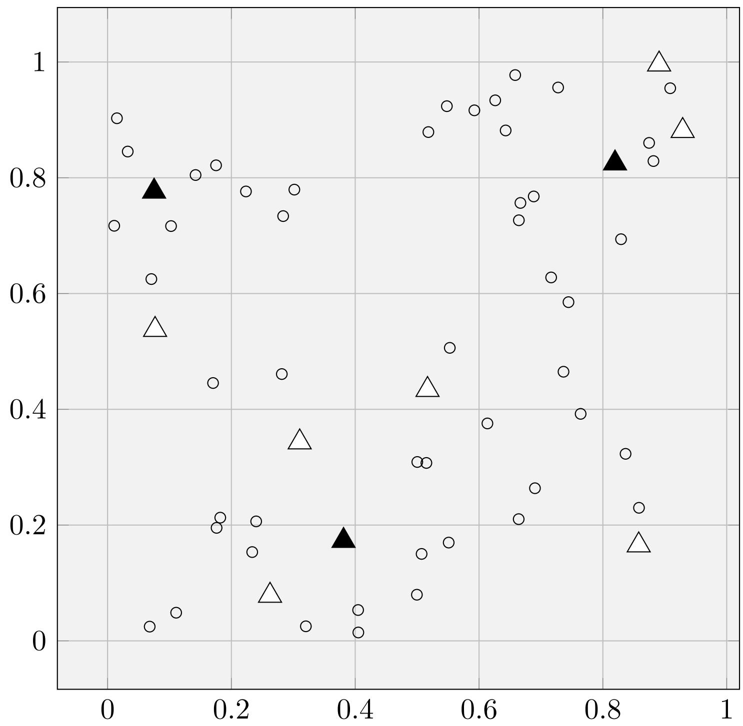


Figure 2.3 – Solution for one facility location instance with $S = 10$ possible sites and $C = 50$ customers on $[0.1]^2$. The uncertainty level is set at $\Gamma = 50\%$. Facility locations are indicated by triangles, customers by open circles. The filled triangles are the locations that are picked to be open.

Table 2.6 – Numerical results for facility location problem with affine policies. The percentages in the last columns depict the optimality gap derived from each upper bound compared to the objective value. All results are averaged over 5 runs.

		Solver time (sec)		Objective value	Upper Bound (Gap%)	
C	Γ%	Primal	Dual		Primal	Combined P/D
10	10	0.2	0.7	30946	32233 (**3.3%**)	32167 (**3.1%**)
	30	0.8	1.2	27894	30474 (**8.0%**)	29835 (**6.1%**)
	50	1.1	1.3	25409	28763 (**10.5%**)	27897 (**7.9%**)
	70	2.0	1.5	23416	24895 (**5.6%**)	24430 (**3.6%**)
	90	2.6	0.9	21889	26511 (**18.3%**)	26353 (**17.5%**)
	100	1.9	0.7	21516	29136 (**28.4%**)	26803 (**19.6%**)
20	10	7.4	3.6	85895	87264 (**1.3%**)	87264 (**1.3%**)
	30	10.4	4.2	79996	82235 (**2.3%**)	81883 (**2.0%**)
	50	18.0	5.2	75404	77060 (**1.8%**)	76827 (**1.6%**)
	70	23.4	5.4	71872	77473 (**6.4%**)	76854 (**5.6%**)
	90	21.2	4.7	69104	69874 (**0.9%**)	69712 (**0.7%**)
	100	11.8	1.1	68226	80301 (**14.7%**)	79810 (**14.1%**)
30	10	55.2	30.3	173069	174547 (**0.7%**)	174004 (**0.5%**)
	30	112.5	35.4	163953	168422 (**2.3%**)	166642 (**1.4%**)
	50	144.3	35.8	156451	160911 (**2.3%**)	157913 (**0.7%**)
	70	220.1	40.8	150070	156881 (**3.6%**)	153511 (**1.8%**)
	90	251.2	31.9	144873	150741 (**3.4%**)	149310 (**2.6%**)
	100	111.8	6.4	143010	164214 (**12.4%**)	159182 (**9.5%**)
40	10	307.4	114.5	243639	244628 (**0.3%**)	244219 (**0.2%**)
	30	787.8	220.7	230556	234272 (**1.3%**)	233557 (**1.1%**)
	50	986.2	197.4	219446	222396 (**1.1%**)	221665 (**0.8%**)
	70	1735.4	199.0	209942	212479 (**1.0%**)	211588 (**0.7%**)
	90	1761.8	154.9	202456	203607 (**0.5%**)	203011 (**0.2%**)
	100	877.7	25.7	200044	223373 (**9.7%**)	222408 (**9.3%**)
50	10	1049.0	326.3	341060	341951 (**0.2%**)	341859 (**0.2%**)
	30	2153.2	530.4	323989	327184 (**0.8%**)	325526 (**0.4%**)
	50	2766.5	557.1	308882	312840 (**1.1%**)	311457 (**0.7%**)
	70	4542.5	536.8	295599	298961 (**1.0%**)	298129 (**0.7%**)
	90	5830.9	469.6	284574	292716 (**2.3%**)	291174 (**1.8%**)
	100	3582.1	68.2	280704	304575 (**7.1%**)	302579 (**6.5%**)

The most striking result is that the dual formulation with affine policies is again solved an order of magnitude faster than the primal formulation with affine policies. This holds especially true for the larger instances and larger values of Γ. We again look at the dimensions of the primal and the dual formulation using Table 2.1 for its dimensions and Table 2.2 for the different constraints. For the the case with $C = 50$ customers we present these results in Table 2.7 and Table 2.8.

Table 2.7 – Comparing dimensions of variables, uncertainties and number of constraints in the primal and dual formulation for the facility location problem (2.20) with $C = 50$ customers.

	Primal formulation (2.2)	Dual formulation (2.3)
# uncertain parameters	50	61
# wait-and-see decisions	500	51
# constraints on variables	61	51
# constraints on uncertain parameter	51	501

Table 2.8 – Comparing the number of affine constraints and sign restrictions in (2.12) and (2.15) for the facility location problem (2.20) with $C = 50$ customers.

	Primal formulation (2.12)	Dual formulation (2.15)
# affine constraints	28661	6222
# sign restrictions	28661	51000

Again we see a smaller number of difficult affine constraints in the dual version in exchange for a larger number of easy-to-handle sign restrictions.

If we take a look at the bounds we see they are very close to the objective value, which shows that the use of affine policies is nearly optimal. This observation was also made for the smaller instances in Ardestani-Jaafari and Delage (2017). For $\Gamma = 100\%$, the lower bound is the most far away from the objective value. This is surprising, as for this case (box uncertainty) we know that affine (in fact, static) policies are provably optimal Ben-Tal et al. (2009, Theorem 14.2.4).

2.7 Concluding remarks

In this chapter, we have used duality for the second-stage decisions and uncertain parameters to derive an equivalent formulation of a primal two-stage adaptive robust optimization model. The resulting dualized formulation is again a two-stage adaptive robust optimization model. We show that optimal affine policies for the primal formulation can be directly constructed from optimal affine policies in the dual formulation. Via two examples of lot-sizing and a facility location problem, we show that the dualized models, when coupled with affine policies, can reduce computational time to solve adaptive problems by an order of magnitude. Furthermore, we provide an algorithm that uses the affine policies in the dual model to strengthen bounds on the optimality gap of affine policies.

CHAPTER 3

Dual approach to two-stage nonlinear robust optimization

3.1 Introduction

Robust optimization is a methodology that can deal with linear and convex optimization models that have parameters that are subject to uncertainty (Ben-Tal and Nemirovski 1998; Ben-Tal and Nemirovski 1999; Ben-Tal and Nemirovski 2002; Ben-Tal et al. 2015; Bertsimas and Sim 2004). In robust optimization all decisions are made here-and-now before the values of the uncertain parameters are known. Adjustable robust optimization is an extension of the robust optimization methodology to handle optimization problems where decisions can be made dynamically over time and additional information about the uncertain parameter is revealed in each stage. In these optimization models one is allowed to have "wait-and-see" decisions that can be decided upon after the true value of the uncertain parameters is known. Since the initial introduction of adjustable robust optimization by Ben-Tal et al. (2004), there has been a wealth of practical problems that have been modeled using linear adjustable robust optimization such as inventory models (Ben-Tal et al. 2004; Ben-Tal et al. 2005), facility location planning (Ardestani-Jaafari and Delage 2017; Atamtürk and Zhang 2007; Gabrel et al. 2014a), energy production scheduling, (Bertsimas et al. 2013; Ng and Sy 2014), project management (Wiesemann et al. 2012), portfolio optimization (Calafiore 2008; Calafiore 2009; Rocha and Kuhn 2012) and capacity expansion planning (Ordóñez and Zhao 2007). Adjustable robust optimization models are in general intractable and NP-hard (Guslitzer 2002). Fortunately, good solutions can be found using linear decision rules. Rather than allowing the wait-and-see decision to depend arbitrarily on the uncertain parameter, linear decision rules restrict the dependence to be affine. The new (here-and-now) decision variables are then the variables in the coefficients of the affine decision rule. In this way, the resulting model is again a linear robust optimization model that can be solved using standard robust optimization techniques, see Ben-Tal et al. (2009). The key benefit is that the model with linear decision rules is of the same optimization class as the static robust version where all decisions have to be made here-and-now. There have been several special cases that show that affine dependence is not a restriction at all, meaning that linear decision rules are optimal for those cases (Bertsimas et al. 2010; Iancu et al. 2013; Gounaris et al. 2013). There are also several other papers that establish optimality or give theoretical a-priori bounds on the objective value (Bertsimas and Goyal 2012; Bertsimas and Bidkhori 2015). Another recently developed method, that is also used in this chapter, is Fourier-Motzkin elimination for adjustable robust optimization (Zhen et al. 2016). This method can solve small adjustable robust linear optimization models to optimality by eliminating the wait-and-see decisions. For larger problems we can eliminate some of the wait-and-see decisions and use linear decision rules for the remaining ones.

Virtually all applications of adjustable robust optimization in the literature have constraints that are linear in the decision variables. This is in sharp contrast to static robust optimization methods where convex nonlinear constraints can be dealt with since the early papers of robust optimization (Ben-Tal and Nemirovski 1998; Ben-Tal and Nemirovski 2002). Static robust optimization nowadays can deal effectively with a large variety of constraints that are convex in the decision variables and concave in the uncertain parameters, see for an overview Ben-Tal et al. (2015). We believe that the main reason behind the lack of papers describing nonlinearities in adjustable robust optimization models lies in the combination of linear decision rules and convexity assumptions that are usually required in robust optimization. To solve static robust models one requires simultaneous convexity in the decision variables and concavity in the uncertain parameter. Suppose we have a problem that is modeled using adjustable robust optimization and happens to be linear in the uncertain parameters, but convex in the wait-and-see decisions. To obtain a static robust model one could try to substitute a linear decision rule for the wait-and-see decisions. However, after substituting the linear decision rule, the model becomes convex in the uncertain parameters. The convexity in the uncertain parameter then prevents us from applying standard robust optimization techniques. Another way to solve these nonlinear adjustable models is to solve the static version of the model. This approach is in general conservative, or even makes the models infeasible, as shown for the linear case in Ben-Tal et al. (2004).

There are only a few papers of nonlinear adjustable robust optimization known to the authors. Pınar and Tütüncü (2005) study a two-period adjustable robust portfolio problem to identify robust arbitrage opportunities when the uncertainty is ellipsoidal. They derive optimal decision rules from exploiting the explicit structure of their formulation, but it is unclear how the generalizations with more constraints, other uncertainty sets or other models would work. Takeda et al. (2008) consider a nonlinear adjustable robust model with polyhedral uncertainty set, similar to the models considered in this chapter. They solve a sampled model, while enumerating all vertices of the polytope uncertainty set. This quickly becomes unviable for even medium sized problems as the number of extreme points of the uncertainty set is exponential in the dimension of the uncertain parameter. Boni and Ben-Tal (2008) consider adjustable robust optimization models with conic quadratic constraints with ellipsoidal uncertainty sets. They approximate the model with linear decision rules using a semidefinite optimization model.

In this chapter we come up with a tractable approach for adjustable robust optimization models that are convex in the wait-and-see decisions. Our method uses the same philosophy as in Chapter 2, which deals with the simpler case where all constraints are linear in the variables. It was shown in that chapter that the dualized model could solve the model several orders of magnitude faster. However, note that in the linear case the original primal version of the adjustable robust optimization models could be solved with linear decision rules as well. In this chapter we consider a more general setting, which allows us to consider models that are nonlinear in the wait-and-see decisions. Our framework can also be used for static robust optimization where constraints are convex in the uncertain parameter. Original static robust optimization models that are convex in the uncertain parameter cannot be reformulated into tractable models using standard robust optimization techniques. We show via explicit examples how these seemingly intractable problems can be formulated as two-stage nonlinear adjustable robust optimization models. Using our dual approach this allows us to find (approximate) solutions to these models via linear decision rules or other methods for linear adjustable robust optimization. Apart from providing the first tractable way of solving the more general nonlinear adjustable robust models, we show how scenarios in the primal and dualized version are tied to each other. This was even for the linear case, considered in Chapter 2, not known before. To summarize, our contributions in this chapter are:

1. We develop an approach in which we consecutively dualize over wait-and-see decisions, with fixed recourse and uncertain parameters in two-stage nonlinear adjustable robust optimization models that have a polyhedral uncertainty set. The resulting model is equivalent to the original one, i.e., the feasible region of the here-and-now decisions and the optimal objective value are the same. Because of the linear structure, all methods for adjustable robust optimization in the literature, such as linear decision rules and Fourier-Motzkin elimination, can be used to find solutions.

2. We show via explicit examples that the dualized version with linear decision rules are of the same optimization class as the static robust versions of the model. For example, the dualized formulation of a two-stage second-order cone model with uncertain parameters, and solved via linear decision rules, results in another tractable second-order cone optimization model.

3. By introducing auxiliary wait-and-see decisions we reformulate some static robust optimization problems that are convex in the uncertain parameter. The resulting models are nonlinear adjustable and fit into our framework. In this way we are able to use linear decision rules to find solutions for a class of robust optimization problems that were deemed intractable before. This class includes models such as robust regression with polyhedral uncertainty sets.

4. Since linear decision rules are in general conservative, we need to provide lower bounds on the optimal objective value. We show how to obtain lower bounds using techniques from Hadjiyiannis et al. (2011). We also show how binding scenarios from the original uncertainty set can be obtained from binding scenarios in the dual formulation.

5. We show that we can use our method to efficiently solve practical two-stage nonlinear robust optimization models and some static robust optimization models that are convex in the uncertain parameters. This is done via two numerical experiments: a commitment model for lot-sizing with distribution on a network and a wireless sensor location problem. We use both Fourier-Motzkin elimination and linear decision rules to find solutions for the dualized formulations. Via the lower bound method we give empirical evidence that linear decision rules give near optimal solutions for our examples.

The rest of this chapter is organized as follows. In Section 3.2 we present our framework and derive our dualized formulation. We also present explicit examples of models that fit into our framework. In Section 3.3 we explain how we obtain lower bounds on the optimal objective value to assess the quality of our solutions. Our numerical examples are presented in respectively Sections 3.4 and 3.5.

Notation. The function g^* is the convex conjugate of the function $g : \mathbb{R}^{n_y} \to \mathbb{R}$ and is defined by

$$g^*(z) = \sup_{y \in \mathrm{dom}(g)} \left\{ y^\top z - g(y) \right\},$$

where $\mathrm{dom}(g)$ is the domain of the function g.

3.2 Linear dual formulation

3.2.1 Framework

We consider the following general two-stage nonlinear robust optimization models:

$$\begin{aligned}
\min_{x\in\mathcal{X},y(\zeta)}\ \max_{\zeta\in\mathcal{U}}\quad & f_0(x)+g_0(y(\zeta)) \\
\text{s.t. } \forall\zeta\in\mathcal{U}\ :\quad & \begin{cases}\zeta^\top F_i(x)+f_i(x)+g_i(y(\zeta))\le 0 & i=1,\dots,m_1\\ A(\zeta)x+By(\zeta)=b(\zeta),\end{cases}
\end{aligned} \tag{3.1}$$

where $\mathcal{X}\subseteq\mathbb{R}^{n_x}$, the functions $g_i:\mathbb{R}^{n_y}\to\mathbb{R}$ are proper closed convex for all $i=0,\dots,m_1$, $F_i(x)=(F_{i,1}(x),\dots,F_{i,n_\zeta}(x))$ and $f_i:\mathbb{R}^{n_x}\to\mathbb{R}$, $F_{i,l}:\mathbb{R}^{n_x}\to\mathbb{R}$ are real valued functions for all $i=0,\dots,m_1$ and $l=1,\dots,n_\zeta$. The matrices $A(\zeta)\in\mathbb{R}^{m_2\times n_x}$ and the vector $b(\zeta)\in\mathbb{R}^{m_2}$ are subject to uncertainty and depend on the uncertain parameter $\zeta\in\mathbb{R}^{n_\zeta}$ in an affine way:

$$A(\zeta)=A^0+\sum_{l=1}^{n_\zeta}A^l\zeta_l,\qquad b(\zeta)=b^0+\sum_{l=1}^{n_\zeta}b^l\zeta_l, \tag{3.2}$$

with $A^l\in\mathbb{R}^{m_2\times n_x}$ and $b^l\in\mathbb{R}^{m_2}$ for all $l=0,\dots,n_\zeta$. We consider the fixed recourse case, meaning that the functions g_i, $i=1,\dots,n$ and the matrix $B\in R^{m_2\times n_y}$ are not subject to uncertainty. Throughout this chapter we focus on polyhedral uncertainty sets of the form

$$\mathcal{U}=\{\zeta\ge 0: D\zeta\le d\}. \tag{3.3}$$

We cannot solve model (3.1) with linear decision rules as is. If we substitute linear decision rules $y(\zeta)=Q\zeta+q$, then the objective and constraints have terms $g_i(Q\zeta+q)$ for all $i=0,\dots,m_1$, which is convex instead of concave in the uncertain parameters if g_i is not linear. Robust optimization techniques such as described in Ben-Tal et al. (2015) require the objective and constraints to be concave in the uncertain parameter as the reformulation maximizes over ζ. Hence, the model seems intractable at first, but via a dual approach we can derive a more tractable formulation.

3.2.2 Duality Theorem

To apply our dualization approach, we require the following property of *strong* relatively complete recourse for our models.

Assumption 1 (Strong relatively complete recourse) *For all $x\in\mathcal{X}$ and all $\zeta\in\mathcal{U}$ there exists a $y\in\bigcap_{i=0}^{m_1} ri(dom(g_i))$, the intersection of the relative interiors of* *the domains of* $g_1, \ldots, g_{m_1}$*, such that*

$$\begin{cases} \zeta^\top F_i(x) + f_i(x) + g_i(y) \leq 0 & i = 1, \ldots, m_1 \\ A(\zeta)x + By = b(\zeta) \end{cases}$$

and for all $i = 1, \ldots, m_1$ *for which* g_i *is nonlinear we have strict feasibility*

$$\zeta^\top F_i(x) + f_i(x) + g_i(y) < 0.$$

This assumption implies that each here-and-now decision is strictly feasible. This assumption is required to guarantee strong duality by Slaters' condition in our dualization procedure. It seems to be very restrictive from a modeling perspective at first. However, in practice models can be cast in such a way that undesirable here-and-now decisions x will result in very high second stage costs $g_0(y(\zeta))$ and the slightly weaker condition of relatively complete recourse (that does not require strict feasibility) is common in two-stage stochastic and robust optimization, see Birge and Louveaux (2011). Another restriction that is imposed by the structure of (3.1) is that the functions g_i and the matrix B do not depend on ζ, which is called the *fixed recourse* case. Loosely speaking, fixed recourse implies that there are no direct interaction terms between ζ and y, such as products $\zeta^\top y$ etc. We do note that the framework in (3.1) is more flexible, e.g., functions $g(x, y, \zeta)$ also fit into the framework for special structures by introducing additional wait-and-see decisions and constraints as shown by examples in this section later.

Most two-stage models are intuitively formatted in the form (3.1) using the "$\forall \zeta \in \mathcal{U}$" notation which is also common in static robust optimization. An equivalent representation of (3.1) is the *min-max-min* formulation

$$\min_{x \in \mathcal{X}} \max_{\zeta \in \mathcal{U}} \min_{y} \left\{ f_0(x) + g_0(y) \middle| \zeta^\top F_i(x) + f_i(x) + g_i(y) \leq 0 \; i = 1, \ldots, m_1, \; A(\zeta)x + By = b(\zeta) \right\}. \tag{3.4}$$

The min-max-min formulation and the formulation in (3.1) are used both in many papers without proof of equivalence, but a formal proof of equivalence is given in Takeda et al. (2008). We use both formulations as the min-max-min formulation proves to be intuitive for our dualization procedure, as the original formulation (3.1) is more intuitive when solving the models via linear decision rules.

We can formulate an equivalent formulation via an approach that we call *consecutive dualization*. This procedure first dualizes over the wait-and-see decision y to obtain a *min-max-max* model and then consecutively dualizes over the uncertain parameter ζ.

This procedure, and the resulting dualized formulation, is described in the following theorem.

Theorem 3.1 *Let $\mathcal{U}$ be a polyhedral set as in* (3.3) *and assume that Assumption 1 holds. Then model* (3.1) *is equivalent to the following dualized model:*

$$\begin{aligned}
&\min_{x\in\mathcal{X},\lambda(\cdot)} \max_{(u,v,w,z)\in\mathcal{V}} \; f_0(x) + \sum_{i=1}^{m_1} v_i f_i(x) + d^\top \lambda(u,v,w,z) + w^\top \left(A^0 x - b^0\right) - \sum_{i=0}^{m_1} z_i \\
&s.t. \; \forall (u,v,w,z)\in\mathcal{V}: \\
&\quad \begin{cases} \sum_{j=1}^{p} D_{j,l}\lambda_j(u,v,w,z) \geq w^\top (A^l x - b^l) + \sum_{i=1}^{m^1} v_i F_{i,l}(x) & l = 1,\dots,n_\zeta \\ \lambda_j(u,v,w,z) \geq 0 \qquad j = 1,\dots,p, \end{cases}
\end{aligned} \tag{3.5}$$

where

$$\mathcal{V} = \left\{ (u,v,w,z):\; v\geq 0,\; v_0 = 1,\; v_i\,(g_i)^*\left(\tfrac{u_i}{v_i}\right) \leq z_i < \infty \;\forall i = 0,\dots,m_1, \right. \\ \left. \sum_{i=0}^{m_1} u_i = -B^\top w \right\}. \tag{3.6}$$

Proof. We use the min-max-min formulation of (3.1), which is given by (3.4), and consider the inner minimization problem over y for a given $x\in\mathcal{X}$ and $\zeta\in\mathcal{U}$. Since Assumption 1 holds we can apply the Fenchel duality Theorem, see Rockafellar (1970), to obtain

$$\min_{x\in\mathcal{X}} \max_{\zeta\in\mathcal{U}} \max_{(u,v,w,z)\in\mathcal{V}} f_0(x) + \sum_{i=1}^{m_1} v_i \left(\zeta^\top F_i(x) + f_i(x)\right) + w^\top \left(A(\zeta)x - b(\zeta)\right) - \sum_{i=0}^{m_1} z_i, \tag{3.7}$$

where $\mathcal{V}$ is as in (3.6). We can then switch the order of maximization such that the inner maximization is over $\zeta\in\mathcal{U}$. Since the inner maximization model is linear in ζ, we can apply strong duality for linear optimization to obtain

$$\min_{x\in\mathcal{X}} \max_{(u,v,w,z)\in\mathcal{V}} \min_{\lambda\geq 0} \left\{ f_0(x) + \sum_{i=1}^{m_1} v_i f_i(x) + d^\top\lambda + w^\top\left(A^0x - b^0\right) - \sum_{i=0}^{m_1} z_i \;\middle|\; \sum_{j=1}^{p} D_{j,l}\lambda_j \geq w^\top (A^l x - b^l) + \sum_{i=1}^{m^1} v_i F_{i,l}(x),\;\; l = 1,\dots,n_\zeta \right\}.$$

This min-max-min formulation is equivalent to (3.5) by the same reasoning that (3.1) is equivalent to the min-max-min formulation (3.4). ■

In the specific linear case, where $F_i(x), f_i(x)$ and $g_i(y)$ are affine functions, Theorem 3.1 coincides with the result in Theorem 2.1 from Chapter 2. For the more general case it is interesting to see that the constraints in the resulting model (3.5) are linear and the uncertainty set $\mathcal{V}$ is convex. The main benefit (and purpose) of dualization is that the resulting model is linear in the wait-and-see decisions and can therefore be solved with any method applicable to linear two-stage models such as linear decision rules (Ben-Tal et al. 2004) and Fourier-Motzkin elimination (Zhen et al. 2016). There are many optimal decision rules that have been characterized for linear two-stage robust models. For instance, it can be shown that there exist polynomials of (at most) degree $3m_1 + m_2$ that are optimal decision rules in (3.5), see Zhen and den Hertog (2017). Furthermore, it is shown in Zhen and den Hertog (2017) that piecewise affine functions are also optimal decision rules for linear adjustable robust optimization problems such as (3.5). More specifically, if $\mathcal{U}$ is simplicial, that paper proves that linear decision rules are optimal; if $\mathcal{U}$ is a box, there are two-piecewise affine functions that are optimal resulting in so-called sum-of-max problems. Techniques proposed in Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016b) can then be applied to find solutions for the sum-of-max problems.

In many cases $\mathcal{V}$ is second-order cone representable, making it a second-order cone (SOC) problem, which can be efficiently solved with off-the-shelve solvers. One interpretation of the dual approach is that the linear structure of the uncertainty set appears in the constraints and the convex structure of the wait-and-see decision is in the new uncertainty set in the dual formulation. Many uncertainty sets naturally require $\zeta \geq 0$, which is the reason we impose it here. However, the nonnegativity restriction on ζ in (3.3) can be omitted. In that case one will end up with equality constraints in the first n_ζ constraints of (3.5). Each equality constraint can be eliminated by eliminating one of the wait-and-see decisions that appear in that equality constraint. Note that we did not need to assume convexity for the functions f_i and F_i in model (3.1) for Theorem 3.1, but to end up with tractable models we usually assume these functions are (componentwise) convex. Finally, we note that one can either apply the result of Theorem 3.1, or use the procedure for consecutive dualization on the problem directly.

3.2.3 Examples of two-stage nonlinear robust models

We give a few examples of two-stage nonlinear robust model formats where Theorem 3.1 can be applied to obtain two-stage linear robust models.

Example 3.1 (Two-stage quadratic robust model) *Consider a two-stage model* *with one quadratic constraint:*

$$\begin{aligned}\min_{x\in\mathcal{X}}\max_{\zeta\in\mathcal{U}}\quad & (c^1)^\top x + (c^2)^\top y(\zeta)\\ \text{s.t.}\qquad & \forall\zeta\in\mathcal{U}: \begin{cases}\frac{1}{2}y(\zeta)^\top Q y(\zeta) + q^\top y(\zeta) + r \le 0\\ A(\zeta)x + By(\zeta) = b(\zeta),\end{cases}\end{aligned} \tag{3.8}$$

where $Q \in \mathbb{R}^{n_y\times n_y}$ is a positive definite matrix, $q \in \mathbb{R}^{n_y}$, $r \in \mathbb{R}$ and $\mathcal{U}$ a polyhedral uncertainty set as in (3.3)*. We assume that the model satisfies Assumption 1. Note that the model could have many more quadratic constraints and also include here-and-now decisions and uncertain parameters in the quadratic constraints. For ease of notation we only consider one quadratic constraint that involves the wait-and-see decision $y(\zeta)$. The model fits the format in* (3.1)*, so we can apply Theorem 3.1 which gives us*

$$\begin{aligned}\min_{x\in\mathcal{X},\lambda(v,w,z)}\ \max_{(v,w,z)\in\mathcal{V}}\quad & (c^1)^\top x + d^\top\lambda(v,w,z) + w^\top\left(A^0x - b^0\right) - z\\ \text{s.t.}\ \ \forall(v,w,z)\in\mathcal{V}: \quad & \begin{cases}\sum_{j=1}^p D_{j,l}\lambda_j(v,w,z) \ge w^\top(A^l x - b^l) & l = 1,\dots,n_\zeta\\ \lambda_j(v,w,z)\ge 0 & j = 1,\dots,p,\end{cases}\end{aligned} \tag{3.9}$$

where $\mathcal{V} = \left\{(v,w,z):\ v\ge 0,\ vr + \frac{1}{v}\left(c^2 + qv - B^\top w\right)^\top Q^{-1}\left(c^2+qv-B^\top w\right) \le z\right\}$.
This uncertainty set $\mathcal{V}$ is second-order cone representable, see Lobo et al. (1998).

The dualized model (3.9) can be solved with linear decision rules, in which case it becomes a conic quadratic optimization model. This is the same optimization class as (3.8) when one takes a single fixed value for ζ (no uncertainty). We encourage users to include *all* auxiliary parameters that arise from the second-order cone representation in the decision rule as this might improve over just using the primitive uncertainties v and w. For more on possible improvements by including auxiliary parameters, see Chapter 4 and Chen and Zhang (2009). The quadratic model in Example 3.1 was already in the correct format of (3.1), which might not always be true for a two-stage robust model. In those cases, we can in some cases still obtain a model that satisfies the format of (3.1) by introducing auxiliary wait-and-see decisions as the next example shows.

Example 3.2 (Two-stage SOC robust model) *Consider a two-stage model with one second-order cone constraint:*

$$\begin{aligned}\min_{x\in\mathcal{X},y(\zeta)}\ \max_{\zeta\in\mathcal{U}}\quad & (c^1)^\top x + (c^2)^\top y(\zeta)\\ \text{s.t.}\qquad & \forall\zeta\in\mathcal{U}:\ \|A(\zeta)x + By(\zeta) - b(\zeta)\|_2 \le (s^1)^\top x + (s^2)^\top y(\zeta) + r,\end{aligned} \tag{3.10}$$

where $c^1, s^1 \in \mathbb{R}^{n_x}$, $c^2, s^2 \in \mathbb{R}^{n_y}$, $r \in \mathbb{R}$ *and a polyhedral uncertainty set* $\mathcal{U}$ *as in* (3.3). *We assume that for every* $x \in \mathcal{X}$ *and* $\zeta \in \mathcal{U}$ *there exists an* y *that strictly satisfies the SOC constraint. The model could have many more SOC or other convex constraints, but for ease of exposition we consider just one SOC constraint. The model is not yet in the format of* (3.1). *However, if we introduce an extra wait-and-see decision* $z(\zeta) \in \mathbb{R}^{m_1}$ *then we see that model* (3.10) *is equivalent to:*

$$\min_{x\in\mathcal{X},y(\zeta),z(\zeta)} \max_{\zeta\in\mathcal{U}} \quad (c^1)^\top x + (c^2)^\top y(\zeta)$$

$$\text{s.t.} \qquad \forall \zeta \in \mathcal{U} : \begin{cases} \|z(\zeta)\|_2 \leq (s^1)^\top x + (s^2)^\top y(\zeta) + r \\ z(\zeta) = A(\zeta)x + By(\zeta) - b(\zeta). \end{cases}$$

By Theorem 3.1, this model is equivalent to

$$\min_{x\in\mathcal{X},\lambda(v,w)} \max_{(v,w)\in\mathcal{V}} \quad (c^1 - vs^1)^\top x + d^\top \lambda(v,w) + w^\top \left(A^0 x - b^0\right)$$

$$\text{s.t. } \forall (w,v) \in \mathcal{V} : \quad \begin{cases} \sum_{j=1}^{p} D_{j,l}\lambda_j(v,w) \geq w^\top(A^l x - b^l) \quad l = 1,\ldots,n_\zeta \\ \lambda_j(v,w) \geq 0 \qquad j = 1,\ldots,p, \end{cases}$$

where $\mathcal{V} = \Big\{(v,w) : \ v \geq 0, \ B^\top w = (c^2) - v(s^2), \ \|w\|_2 \leq v\Big\}$.

Example 3.2 shows that sometimes one has to introduce auxiliary wait-and-see decisions in order to obtain a model that fits the format of model (3.1). Note that in general we can have many types of substitions. For example, if $\tilde{A}(\zeta)$ and $\tilde{b}(\zeta)$ are as in (3.2) and $g : \mathbb{R}^{m_2} \to \mathbb{R}$ is a proper closed convex function, then a constraint of the form

$$\zeta^\top F(x) + f(x) + g\left(\tilde{A}(\zeta)x + By(\zeta) - \tilde{b}(\zeta)\right) \leq 0$$

can be replaced by the following system of inequalities:

$$\begin{aligned} \zeta^\top F(x) + f(x) + g(z(\zeta)) &\leq 0 \\ \tilde{A}(\zeta)x + By(\zeta) - \tilde{b}(\zeta) &= z(\zeta), \end{aligned}$$

where $z(\zeta) \in \mathbb{R}^{m_2}$ is an additional wait-and-see variable. Another example is when $g : \mathbb{R}^{m_2} \to \mathbb{R}$ and $\tilde{g}_i : \mathbb{R}^{m_2} \to \mathbb{R}$ are proper closed convex functions, with g nondecreasing. In that case, a constraint of the form

$$\zeta^\top F(x) + f(x) + g(h(\zeta, x, y(\zeta))) \leq 0,$$

where $h_k(\zeta, x, y(\zeta)) = \zeta^\top \tilde{F}_k(x) + \tilde{f}_k(x) + \tilde{g}_k(y(\zeta))$ for all $k = 1,\ldots,m_3$, can be replaced by the following system of inequalities:

$$\begin{aligned} \zeta^\top F(x) + f(x) + g(z(\zeta)) &\leq 0 \\ \zeta^\top \tilde{F}_k(x) + \tilde{f}_k(x) + \tilde{g}_k(y(\zeta)) &\leq z_k(\zeta) \qquad k = 1,\ldots,m_3, \end{aligned}$$

where $z(\zeta) \in \mathbb{R}^{m_3}$ is again an additional wait-and-see variable. All these substitutions are made to ensure that the resulting system of inequalities fits into the format of model (3.1).

3.2.4 Examples of static robust models that are convex in uncertain parameters

Theorem 3.1 is designed for two-stage problems and does not seem to be fit for static problems at first sight. However, by using auxiliary wait-and-see decisions, we can use it for some static robust models with constraints that are convex in the uncertain parameters. These models are, to the best of the authors' knowledge, deemed intractable in general as robust optimization techniques require constraints to be concave in the uncertain parameter, see Ben-Tal et al. (2015).

Example 3.3 (Static robust model with objective convex in ζ) *Consider the robust optimization model*

$$\min_{x\in\mathcal{X}}\max_{\zeta\in\mathcal{U}} \quad g(A(\zeta)x - b(\zeta))$$

where $g:\mathbb{R}^{n_y}\to\mathbb{R}$ is a proper closed convex function and $\mathcal{U}$ a polyhedral uncertainty set as in (3.3)*. The current model is convex in ζ in the objective function and can therefore not be reformulated using standard techniques in robust optimization. However, we can introduce additional wait-and-see decisions $y(\zeta)$ to obtain*

$$\begin{aligned}\min_{x\in\mathcal{X}}\max_{\zeta\in\mathcal{U}} \quad & g(y(\zeta))\\ s.t. \quad & \forall\zeta\in\mathcal{U}: y(\zeta) = A(\zeta)x - b(\zeta).\end{aligned}$$

This model satisfies the format of (3.1)*, so by Theorem 3.1 this model is equivalent to*

$$\begin{aligned}\min_{x\in\mathcal{X},\lambda(w,z)}\ \max_{(w,z)\in\mathcal{V}} \quad & d^\top\lambda(w,z) + w^\top\left(A^0x - b^0\right) - z\\ s.t.\ \forall(w,z)\in\mathcal{V}: \quad & \begin{cases}\sum_{j=1}^p D_{j,l}\lambda_j(w,z) \geq w^\top(A^lx - b^l) & l = 1,\ldots,n_\zeta\\ \lambda(w,z)\geq 0,\end{cases}\end{aligned}$$

where $\mathcal{V} = \{(w,z):\ g^*(w)\leq z<\infty\}$.

The model from Example 3.3 includes models such as robust regression (with $g(y) = \|y\|_2$). These models have exact tractable formulations for some specific seminorm uncertainty sets which results in well-known regularized regression models such as LASSO or regression with Tikhonov regularization, see El Ghaoui and Lebret (1997), Xu et al. (2009) and Bertsimas and Copenhaver (2017). For other popular uncertainty sets, such as the budget uncertainty set by Bertsimas and Sim (2004), no exact tractable reformulation is known. With our dual approach we give a way to (approximately) solve these models with these uncertainty sets using any method developed for linear two-stage robust models. We must note that the procedure is not useful in an optimization setting if the model contains constraints of the form $f(x,\zeta)\leq 0$

that are convex in ζ. Assumption 1 requires that each here-and-now decision x and uncertain parameter $\zeta \in \mathcal{U}$ is feasible. Abiding by that assumption we would have that the constraint $f(x,\zeta) \leq 0$ is satisfied for all $x \in \mathcal{X}$ and $\zeta \in \mathcal{U}$, which means that the constraint would a-priori be known to be redundant. In cases with convex uncertainty in the constraints it is therefore likely that Assumption 1 does not hold. However, we can still model feasibility problems with constraints that are convex in the uncertain parameter ζ as the next example shows.

Example 3.4 *Consider a general feasibility model*

$$\begin{aligned} &\min_{x\in\mathcal{X}} \ 0 \\ &s.t. \quad \forall \zeta \in \mathcal{U}: \ h_i(\zeta, x) \leq 0 \qquad i = 1,\ldots,m_1, \end{aligned}$$

which takes the value 0 *if there exists a feasible* x *and* ∞ *otherwise. The functions* $h_i(\zeta,x)$ *are jointly convex in* (ζ,x) *for all* $i = 1,\ldots,m_1$*, and* $\mathcal{U}$ *is a polyhedral uncertainty set of the form* (3.3)*. This model is not in the format of* (3.1)*, but it can be reformulated to fit the format by using auxiliary wait-and-see variables. We introduce wait-and-see decisions* $y(\zeta)$ *that represent the constraint violation and* $\phi(\zeta)$ *to represent the combined vector* (ζ,x)*:*

$$\begin{aligned} &\min_{x\in\mathcal{X},y(\zeta),\phi(\zeta)} \ \max_{\zeta\in\mathcal{U}} \ \sum_{i=1}^{m_1} y_i(\zeta) \\ &s.t. \qquad \forall \zeta \in \mathcal{U} \ : \begin{cases} h_i(\phi(\zeta)) - y_i(\zeta) \leq 0 & i = 1,\ldots,m_1 \\ y_i(\zeta) \geq 0 & i = 1,\ldots,m_1 \\ \phi(\zeta) = (\zeta,x). \end{cases} \end{aligned}$$

Note that the equality constraints are linear and the functions h_i *are convex in* $z(\zeta)$*. Hence, this model fits the format given in* (3.1)*. Moreover, Assumption 1 is satisfied for all* $x \in \mathcal{X}$ *and* $\zeta \in \mathcal{U}$ *by taking* $\phi(\zeta) = (\zeta,x)$ *and* $y_i(\zeta) = \max\{h_i(z(\zeta)),0\} + 1$ *for all* $i = 1,\ldots,m_1$*. By Theorem 3.1 this model is equivalent to*

$$\begin{aligned} &\min_{x\in\mathcal{X},\lambda(u,v,w,z)} \ \max_{(u,v,w,z)\in\mathcal{V}} \ d^\top \lambda(u,v,w,z) + w^\top x - \sum_{i=0}^{m_1} z_i \\ &s.t. \ \forall (u,v,w,z) \in \mathcal{V}: \begin{cases} \sum_{j=1}^{p} D_{j,l}\lambda_j(u,v,w,z) \geq w_l & l = 1,\ldots,n_\zeta \\ \lambda_j(u,v,w,z) \geq 0 & j = 1,\ldots,p, \end{cases} \end{aligned}$$

where

$$\mathcal{V} = \left\{ (u,v,w,z): \ 0 \leq v \leq 1, \ v_i (h_i)^* \left(\tfrac{u_i}{v_i}\right) \leq z_i < \infty \ \forall i = 0,\ldots,m_1, \ \sum_{i=0}^{m_1} u_i = w \right\}.$$

When using linear decision rules in this dualized formulation, one obtains conservative solutions. For this example the conservativeness implies that whenever the *objective value with linear decision rules is nonpositive a feasible solution exists. If the objective value is positive then the result is inconclusive: either the problem is indeed infeasible, or the linear decision rules are just too conservative. The conservativeness could be further assessed in this case by using the lower bound method in Section 3.3. If the lower bound is also positive, then one can still conclude that the problem is infeasible.*

3.2.5 Challenges with generalizations of model formulation

The tractability of the reformulation of the model (3.1) hinges on the initial structure of the nonlinear model, the uncertainty set and the assumption of relatively complete recourse. In this section we briefly describe the challenges that any changes in this structure or assumption might bring.

If the relatively complete recourse assumption is dropped, then we cannot guarantee strong duality for the convex case. We can still apply duality and obtain the dualized model (3.5), but the objective value might only be a lower bound. Therefore, even if we are able to solve the dualized model to optimality, the resulting here-and-now decision might be infeasible in the original primal formulation. The same holds for the implicit assumption that decision rules for $y(\zeta)$ are continuous. If the original model requires $y(\zeta)$ to be binary, then dualizing the model will result in a lower bound. Nevertheless, we can deal with integer here-and-now variables x as we do not dualize over x. However, the resulting model with integer here-and-now decisions will be a mixed-integer convex optimization model. These models are still solvable for moderate sizes in some cases (such as mixed-integer second-order cone models), but quickly become difficult to solve for larger models.

The other assumptions are more of a structural nature. If the functions $g_i(y(\zeta))$ also depend on either x or ζ (which is the so-called non-fixed recourse case) then one runs into trouble as ζ or x appears in the dualized uncertainty set after one dualization step. However, in some cases we can obtain tractable formulations by introducing auxiliary wait-and-see decisions as was done in Examples 3.2-3.4. For other cases, such as the non-fixed recourse case, there is not much hope to find solutions in an efficient way, as this is already difficult in the simpler linear case that was considered in the seminal paper by Ben-Tal et al. (2004). The final structural assumption that one might relax is the linearity in ζ. By doing so we still obtain a dualized model with strong duality, but the resulting model introduces new wait-and-see decisions which are nonlinear in the wait-and-see decisions. In that case, the benefits of the dual formulations are not clear, as we still have the same difficulties that also arose in (3.1) because of the nonlinearity in the wait-and-see decisions.

Throughout this chapter the focus has been on polyhedral uncertainty sets. In principle the same procedure for consecutive dualization can be applied for uncertainty sets that are not polyhedral. However, the resulting dualized formulation does not become tractable. For instance, if we consider popular ellipsoidal uncertainty sets $\mathcal{U} = \{\zeta : ||\zeta||_2 \leq 1\}$, then the robust constraints in (3.1) require maximization of a norm over the polyhedron $\mathcal{W}$ which is very difficult.

3.3 Bounds on the optimal value

The dualized model (3.5) is linear in the wait-and-see decisions, so good solutions can be found using methods such as linear decision rules, possibly combined with Fourier-Motzkin elimination. These methods are not exact, so the solutions might be suboptimal. It is therefore important to find lower bounds on the optimal objective value of the original model (3.1) to assess the quality of the solutions.

3.3.1 Sampled scenarios

One simple way of obtaining a lower bound is to consider a finite subset $\{\zeta^1, \ldots, \zeta^K\}$ of scenarios from the uncertainty set $\mathcal{U}$. Instead of making a decision rule $y(\zeta)$ that is feasible for all values of $\zeta \in \mathcal{U}$, we only require feasibility for the finite subset to obtain a lower bound. In that case we can attach a single optimization variable y^k to each scenario ζ^k, for $k = 1, \ldots, K$. The lower bound model is therefore the "sampled version" of the original model:

$$\begin{aligned}
\min_{\sigma, x \in \mathcal{X}, y^1, \ldots, y^K} \quad & \sigma \\
\text{s.t.} \quad & f_0(x) + g_0(y^k) \leq \sigma \qquad \forall k = 1, \ldots, K \\
& (\zeta^k)^\top F_i(x) + f_i(x) + g_i(y^k) \leq 0 \qquad \forall i = 1, \ldots, m_1, \ k = 1, \ldots, K \\
& A(\zeta^k) + By^k = b(\zeta^k) \qquad \forall k = 1, \ldots, K.
\end{aligned} \tag{3.11}$$

Model (3.11) is a standard convex optimization model as we do not have robust constraints with '$\forall \zeta \in \mathcal{U}$' in the model anymore. Clearly this is a lower bound, since the solution is only feasible for a finite subset of the uncertainty set. There could be realizations in $\mathcal{U}$ for which a higher objective value is attained, making the here-and-now decision suboptimal. This sampled approach can be applied to any two-stage model, and in particular also to our dualized model (3.5). In the dualized model we would take a finite subset $\{(u^1, w^1, v^1, z^1), \ldots, (u^K, w^K, v^K, z^K)\}$ from $\mathcal{V}$ with a single optimization variable λ^k for each scenario (u^k, w^k, v^k, z^k), $k = 1, \ldots, K$. The sampled version of the dualized model is:

$$\begin{aligned} \min_{\sigma, x\in\mathcal{X}, \lambda^1,\ldots,\lambda^K\geq 0} \quad & \sigma \\ \text{s.t.} \quad & f_0(x) + \sum_{i=1}^{m_1} v_i^k f_i(x) + d^\top \lambda^k + (w^k)^\top \left(A^0 x - b^0\right) - \sum_{i=0}^{m_1} z_i^k \leq \sigma \\ & \qquad \forall k = 1,\ldots,K \\ & \sum_{j=1}^{p} D_{j,l}\lambda_j^k \geq (w^k)^\top (A^l x - b^l) + \sum_{i=1}^{m^1} (v_i^k) F_{i,l}(x) \\ & \qquad \forall l = 1,\ldots,n_\zeta,\ k = 1,\ldots,K, \end{aligned} \tag{3.12}$$

which is again a standard convex optimization model.

3.3.2 Choosing a good set of scenarios

The question that remains for the sampled model is how to choose the finite set of scenarios. One way to do this would be to include all extreme points from $\mathcal{U}$. In that case, one can prove that the lower bound model is optimal. The proof is similar to the proof for the fully linear case, see Bemporad et al. (2003), but given here for completeness.

Theorem 3.2 *Let $\mathcal{U}$ be a polyhedral uncertainty set with K extreme points $\zeta^1,\ldots,\zeta^K$. Then the optimal here-and-now solution $\bar{x}$ of model* (3.11) *is also optimal for model* (3.1) *and their optimal objective values coincide.*

Proof. Let $\bar{\sigma}, \bar{x}, \bar{y}^1, \ldots, \bar{y}^K$ be the optimal solution of (3.11). We know that the optimal value $\bar{\sigma}$ of the sampled model (3.11) gives a lower bound of (3.1), so it is sufficient to show that $\bar{x}$ is feasible and we can construct a feasible decision rule $y(\zeta)$ that gives an objective value of at most $\bar{\sigma}$. Let $\zeta \in \mathcal{U}$ and write it as the convex combination of the extreme points of $\mathcal{U}$:

$$\zeta = \sum_{k=1}^{K} \alpha_k \zeta^k \tag{3.13}$$

for some $\alpha_1, \ldots, \alpha_K \in [0,1]$, $\sum_{k=1}^K \alpha_k = 1$. We take for the wait-and-see decision $y(\zeta)$ the following value

$$y(\zeta) = \sum_{k=1}^{K} \alpha_k \bar{y}^k, \tag{3.14}$$

with $\alpha_1, \ldots, \alpha_K$ the same values as those in the convex combination of (3.13). Then we have:

$$\begin{aligned}\zeta^\top F_i(\bar{x}) + f_i(\bar{x}) + g_i(y(\zeta)) &= (\sum_{k=1}^K \alpha_k \zeta^k)^\top F_i(\bar{x}) + f_i(\bar{x}) + g_i(\sum_{k=1}^K \alpha_k \bar{y}^k) \\ &\leq \sum_{k=1}^K \alpha_k \left((\zeta^k)^\top F_i(\bar{x}) + f_i(\bar{x}) + g_i(y^k) \right) \\ &\leq 0,\end{aligned}$$

where the first inequality is due to convexity of the functions g_i and the last inequality is due to the fact that $\bar{x}, \bar{y}^1, \ldots, \bar{y}^K$ is feasible for (3.11). Analogously, we can show that for $\bar{x}$ and decision rule $y(\zeta)$ from (3.14) we have $f_0(x) + g_0(y(\zeta)) \leq \bar{\sigma}$ for all $\zeta \in \mathcal{U}$. Hence, the optimal objective value of (3.1) is at most $\bar{\sigma}$. ■

Of course, the set of extreme points of a polyhedral uncertainty set $\mathcal{U}$ is in practice way too large. As we see in our numerical examples, this is most likely only doable when the uncertainty set is low-dimensional. Another way to obtain a small and effective finite set of scenarios for two-stage linear models is described by Hadjiyiannis et al. (2011). That method takes scenarios that are binding for the model solved with linear decision rules, hoping that the same set of scenarios is also binding for the optimal (nonlinear) decision rule. Since it obtains binding scenarios for each constraint, the set of binding scenarios is at most the number of constraints in the model and possibly smaller if some of the scenarios coincide. For more details on the method we refer to the original paper by Hadjiyiannis et al. (2011). One needs to be able to solve the model with linear decision rules to obtain a set of scenarios by the method proposed by Hadjiyiannis et al. (2011). Hence, we can only apply their method to obtain a set of scenarios $\{(u^1, w^1, v^1, z^1), \ldots, (u^K, w^K, v^K, z^K)\}$ for the dualized model because it is linear in the wait-and-see decisions. In Chapter 2 the scenarios are obtained after both the primal and dualized solutions with linear decision rules are obtained. One can solve the original primal model with linear decision rules in that case because the model is fully linear.

3.3.3 Primal scenarios corresponding to dual scenarios

We can establish a link between the primal scenarios $\{\zeta^1, \ldots, \zeta^K\}$ from the original model and the dual scenarios $\{(u^1, w^1, v^1, z^1), \ldots, (u^K, w^K, v^K, z^K)\}$ by using a dual approach. By dualizing over $\lambda_1, \ldots, \lambda_K$ we get the following equivalent formulation of (3.12):

$$\min_{x \in \mathcal{X}} \max_{\zeta \in \mathcal{U}} \max_{1 \leq k \leq K} f_0(x) + \sum_{i=1}^{m_1} v_i^k \left(\zeta^\top F_i(x) + f_i(x) \right) + \left(A(\zeta)x - b(\zeta) \right)^\top w^k - \sum_{i=0}^{m_1} z_i^k, \tag{3.15}$$

which is similar to (3.7), but with the inner maximization over $(u, w, v, z) \in \mathcal{V}$ replaced by the finite subset with K scenarios. For a fixed x we can obtain primal scenarios ζ^k for each k as the maximizers of model (3.15):

$$\zeta^k \in \arg\max_{\zeta \in \mathcal{U}} \left\{ \sum_{i=1}^{m_1} v_i^k \left(\zeta^\top F_i(x) \right) + \left(w^k \right)^\top (A(\zeta)x - b(\zeta)) \right\}. \tag{3.16}$$

The resulting set of scenarios $\{\zeta^1, \ldots, \zeta^K\}$ can then be used in the sampled model (3.11). One can now solve either the primal sampled model (3.11), which is a convex optimization model, or the dual sampled model (3.12), which is a linear model. The latter is much easier to solve since it is a linear model. In general, we cannot know beforehand whether (3.11) or (3.12) gives a stronger lower bound. However, we can always combine the constraints from these sampled models. The resulting model has a smaller feasible region than both individual models and must therefore lead to the tightest lower bound. In case there is only *right-hand-side* uncertainty in model (3.1), and the scenarios have been obtained by (3.16), then we can show that the lower bound from (3.11) is always higher (or equal to) (3.12). We say that there is only *right-hand-side* uncertainty if there is no direct interaction between the here-and-now decisions x and ζ. The more formal definition is given below.

Definition 3.1 (Right-hand-side uncertainty) *Model* (3.1) *has right-hand-side uncertainty if $F_i = 0$ for all $i = 1, \ldots, m_1$ and there exists $\bar{A} \in \mathbb{R}^{m_2 \times n_x}$ such that for all $\zeta \in \mathcal{U}$ we have $A(\zeta) = \bar{A}$.*

Note that in the case of right-hand-side uncertainty, the scenarios ζ^k can be obtained in (3.16) independent of the here-and-now decision x as the only terms depending on ζ are $(w^k)^\top b(\zeta)$.

Theorem 3.3 *Let $\{(u^1, w^1, v^1, z^1), \ldots, (u^K, w^K, v^K, z^K)\}$ be a finite set of dual scenarios and $\{\zeta^1, \ldots, \zeta^K\}$ be a set of primal scenarios obtained from* (3.16)*. If there is only right-hand-side uncertainty in model* (3.1)*, then the lower bound from* (3.11) *is at least as tight as the lower bound from* (3.12)*.*

Proof. By duality for linear programming, (3.12) is equivalent to (3.15). The latter formulation can be written as

$$\min_{x \in \mathcal{X}} \max_{k \in \{1,\ldots,K\}} \left\{ f_0(x) + \sum_{i=1}^{m_1} v_i^k f_i(x) + \left(w^k \right)^\top \left(\bar{A}x - b(\zeta^k) \right) - \sum_{i=0}^{m_1} z_i^k \right\}, \tag{3.17}$$

where ζ^k are the primal scenarios obtained by (3.16). Since (u^k, w^k, v^k, z^k) are in $\mathcal{V}$ for all $k = 1, \ldots, K$, the value of (3.17) must be smaller than or equal to

$$\min_{x \in \mathcal{X}} \max_{k} \max_{(u^k, w^k, v^k, z^k) \in \mathcal{V}} \left\{ f_0(x) + \sum_{i=1}^{m_1} v_i^k \left(f_i(x) \right) + \left(w^k \right)^\top \left(\bar{A}x - b(\zeta^k) \right) - \sum_{i=0}^{m_1} z_i^k \right\},$$

since we maximize over (u^k, w^k, v^k, z^k) in $\mathcal{V}$, instead of a fixing these K values beforehand. The value of this optimization problem is, by dualizing over (u^k, w^k, v^k, z^k), equivalent to (3.11). Hence, the optimal objective value of the model (3.11) is at least as high as the optimal objective value of (3.12). ■

We emphasize that the right-hand-side uncertainty definition is stated for models that are of the format (3.1). This means that it could also apply to models where auxiliary wait-and-see decisions are introduced and right-hand side uncertainty is only visible in the final formulation of the model.

3.4 Example 1: distribution on a network with commitments

This problem is adapted from Section 2.5. For the distribution on a network we determine the stock allocation x_i for location i, and the contracted transporting units z_{ij} from location i to location j, $i, j = 1, ..., N$, prior to knowing the realization of the demand at each location. The demand ζ is uncertain and assumed to be in a budget uncertainty set:

$$\mathcal{U} = \left\{ \zeta \geq 0 \ : \ \zeta \leq \hat{\zeta} e, \ e^\top \zeta \leq \Gamma \right\},$$

the same budget uncertainty set as used in (2.18) in Chapter 2. After we observe the realization of the demand we can transport stock y_{ij} from location i to location j at cost t_{ij} in order to meet all demand, $i, j = 1, ..., N$. The aim is to minimize the worst case total costs, which includes the storage costs (with unit costs c_i), the cost arising from shifting the products from one location to another (after the demands are realized), and the cost from violating the committed contract. A contract is violated if the transporting units y_{ij} differentiate from the committed units z_{ij}, $i, j = 1, ...N$. This distribution model can now be written as a specific instance of the primal problem as follows:

$$\begin{aligned}
\min_{x,z,y(\zeta)\geq 0} \ \max_{\zeta \in \mathcal{U}} \quad & c^\top x + \sum_{i,j=1}^{N} t_{ij} y_{ij} + \frac{1}{2} \sum_{i,j=1}^{N} t_{ij} (y_{ij} - z_{ij})^2 & \\
\text{s.t.} \quad & \sum_{j=1}^{N} y_{ji} - \sum_{j=1}^{N} y_{ij} \geq \zeta_i - x_i & \forall \zeta \in \mathcal{U}, \ i = 1, ..., N \\
& x_i \leq K_i & i = 1, ..., N \\
& \sum_{i=1}^{N} x_i \geq \Gamma, &
\end{aligned} \tag{3.18}$$

where the third term in the objective of (3.18) captures the cost of contract violation. Without this third term in the objective, the model is equivalent to the linear adjustable robust model (2.19) from Chapter 2. The second line contains the balance equations: we have to shift stock to and from location i such that the initial storage plus the net shift in stock still exceeds the demand at i. The constraints in the third line restrict the capacity of the stock at each location. The dualized formulation we obtain after consecutive dualization over the wait-and-see decisions y and the uncertain parameters ζ is given below:

$$\begin{array}{lll} \min\limits_{\substack{x,z\geq 0,\tau \\ \lambda(\cdot)\geq 0}} & c^\top x+\tau & \\ \text{s.t.} & \sum_{i=1}^{N}(\hat{\zeta}\lambda_i - u_i x_i) + \Gamma\lambda_0 - \sum_{i,j=1}^{N}\left[(u_j - u_i - t_{ij} - v_{ij})z_{ij} + \frac{1}{2}w_{ij}\right] \leq \tau & \\ & & \forall (u,v,w)\in\mathcal{W} \\ & \lambda_0 + \lambda_i \geq u_i \quad i=1,...,N & \\ & & \forall (u,v,w)\in\mathcal{W} \\ & x_i \leq K_i \qquad i=1,...,N & \\ & \sum_{i=1}^{N} x_i \geq \Gamma, & \end{array} \tag{3.19}$$

where

$$\mathcal{W} = \left\{(u,v,w)\geq 0: \ (u_i - u_j + v_{ij} - t_{ij})^2 \leq w_{ij}t_{ij}, \quad \forall i,j=1,...N\right\}.$$

Now we can apply Fourier-Motzkin elimination and linear decision rules to solve (3.19).

3.4.1 Numerical setting

We choose $N \in \{5, 10, 20, 30\}$ locations uniformly at random from $[0, 10]^2$. Let t_{ij}, the cost to transport one unit of demand from location i to j, be the Euclidean distance. The unit storage cost c_i are equal to 6 for $i = 1, ..., \lceil N/10 \rceil + 1$ warehouses and 10 for $i = \lceil N/10 \rceil + 1, ..., N$ stores. The individual maximum demand ζ and the capacity K_i, $i = 1, ..., N$, of each location is set to 30 units. The total demand in the network is set to $20\sqrt{N}$. As an illustration, Figure 3.1 depicts a distribution on a network obtained from solving (3.19) with linear decision rules, which takes around 100s. All computations were carried out with MOSEK 8.0 (MOSEK ApS. 2017) on an Intel Core(TM) i5-4590 Windows computer running at 3.30GHz with 8GB of RAM. All modeling was done using the modeling package XProg (http://xprog.weebly.com).

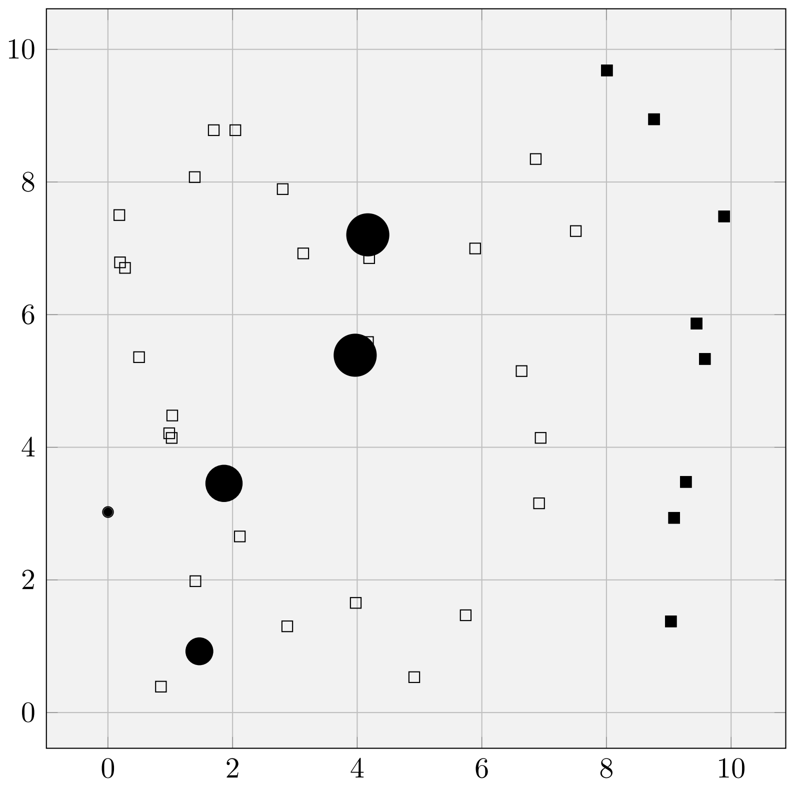


Figure 3.1 – Stock allocation for $N = 40$ with 35 stores (squares) and 5 warehouses (circles) for one random instance. The filled dots have stock and the larger the dots are, the more stock is allocated.

3.4.2 Results

We first consider a small instance and present the results in Table 3.1. One can observe that the solutions converge to optimality as more adjustable variables in (3.19) are eliminated via Fourier-Motzkin elimination. If all $N + 1(= 6)$ adjustable variables are eliminated, the optimal solution can be obtained. Note that Fourier-Motzkin elimination cannot be applied to (3.18) because the adjustable variables appear nonlinearly in the model. By solving (3.18) and (3.19) with static decision rules, we obtain the respective P-S and D-S solutions. For #Elim.= 0, the P-S solutions are far from optimal on average, and the results for P-S and D-S are different, which indicates that the models (3.18) and (3.19) with static decision rules are not equivalent in general. The D-L solutions are obtained by solving the model with linear decision rules in the dual formulation. They perform significantly better than the P-S solutions, the solution of the static robust version of the original model. Since the problem (3.18) has right-hand-side uncertainty, the LB-P lower bounds obtained from the primal scenarios are indeed tighter than the LB-D bound from the dual scenarios (see Theorem 3.3). Hence, we only focus on the LB-P lower bounds for the rest of this chapter.

Table 3.1 – Lot-sizing problem with $N = 5$. #Elim. denotes the number of adjustable variables that are eliminated. P-S and D-S are obtained from solving (3.18) and (3.19) with static decision rules, respectively. D-L is obtained from solving (3.19) with linear decision rules. LB-P and LB-D denote the lower bounds obtained from the primal scenarios (see Section 3.3.3) and the (dual) binding scenarios of Hadjiyiannis et al. (2011), respectively. INF means infeasible. N.A. represents not applicable. All the numbers are the average of 10 randomly generated instances.

#Elim.	0	1	2	3	4	5	6
#Constr.	12	11	11	13	19	33	272
P-S	840	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
D-S	INF	INF	INF	INF	INF	840	607
D-L	677	677	670	656	638	624	607
Time(s)	0.03	0.03	0.03	0.04	0.07	0.15	0.66
LB-P	605	605	606	607	607	607	607
LB-D	8	119	232	361	583	597	607

Table 3.2 considers medium size instances. Due to the 1 hour computational limit, the effectiveness of Fourier-Motzkin elimination diminishes as the problem size becomes larger. Via vertex enumeration (see Theorem 3.2), we obtain the optimal solutions for $N = 10$, and the average optimal objective value is 937. Therefore, the LB-P lower bounds are very tight. When $N = 20$, the vertices of the budget uncertainty set are too many to enumerate, i.e., 83,716 vertices. For #Elim. = 0, the average P-S values are much larger than the average D-L values.

For large instance, using Fourier-Motzkin elimination becomes too time consuming. Hence, we only report the results without using Fourier-Motzkin elimination in Table 3.3. On average, the difference between the values from P-S and D-L becomes much larger as N increases. However, the differences between the LB-P lower bound and the D-L upper bound do not increase as the problem size becomes larger, so the linear decision rules remain near optimal.

3.5 Example 2: sensor network model

Here we consider a problem where there are N points in $\mathbb{R}^2$ that must be connected by links. Some of the N points are already placed and the decision maker has to decide where to place the remaining points. The goal is to minimize the total distance of all the links together. An application would be where the points represent wireless sensors and modules on a network that are interconnected and one wants to minimize

Table 3.2 – Lot-sizing problem with $N \in \{10, 20\}$. #Elim. denotes the number of adjustable variables that are eliminated. P-S is obtained from solving (3.18) with static decision rules. D-L is obtained from solving (3.19) with linear decision rules. LB-P denotes the lower bounds obtained from the primal scenarios (see Section 3.3.3). Time(s) reports the computation time (in seconds) for solving D-L. * means the computation time needed exceeds 1 hour. All the numbers are the average of 10 randomly generated instances.

	#Elim.	0	1	2	3	4	5	6	7	8	9	10	11
N=10	#Constr.	22	21	21	23	29	43	73	135	261	515	1025	149424
	P-S	1840	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	D-L	1029	1029	1028	1021	1014	1006	996	983	971	956	944	*
	LB-P	935	935	936	936	936	936	937	937	937	937	937	*
	Time(s)	0.3	0.4	0.5	0.5	1	1	3	4	10	14	26	*
N=20	#Constr.	42	41	41	43	49	63	93	165	281	535	1045	2067
	P-S	3760	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	D-L	1377	1377	1377	1376	1374	1371	1368	1363	1359	1355	1350	*
	LB-P	1272	1273	1273	1273	1274	1274	1274	1275	1276	1276	1276	*
	Time(s)	14	13	6	9	11	28	44	171	624	1156	2827	*

Table 3.3 – Lot-sizing problem with $N \in \{30, 40, 50, 60\}$. P-S is obtained from solving (3.18) with static decision rules. D-L is obtained from solving (3.19) with linear decision rules. LB-P denotes the lower bounds obtained from the primal scenarios (see Section 3.3.3). Time(s) reports the computation time (in seconds) for solving D-L. All the numbers are the average of 10 randomly generated instances.

N	30	40	50	60
#Constr.	62	82	102	122
P-S	5680	7600	9520	11440
D-L	1606	1790	1962	2115
LB-P	1495	1681	1856	2004
Time(s)	31	118	337	665

the total energy needed for the wireless transmission over the links. This example is based on the placement and location problem, for which the nominal case is described in Boyd and Vandenberghe (2004, Section 8.7). In the nominal case the model can be written as

$$\begin{aligned} \min_{y} \quad & \sum_{(i,j)\in\mathcal{A}} \|y_i - y_j\|_2 \\ \text{s.t.} \quad & y_i = \bar{a}_i \qquad \forall i \in L, \end{aligned}$$

where $y_i \in \mathbb{R}^2$ are the locations of the points for all $i = 1, \dots, N$. The points that are fixed (already placed) are given by the set L and their locations by $\bar{a}_i \in \mathbb{R}^2$. The remaining points not in L are free to set by the optimizer. The set of prescribed (undirected) links is given by $\mathcal{A}$.

3.5.1 The model

Here we consider the case where the locations of the fixed sensors $\bar{a}_i$ are not precisely known. This uncertainty in the locations could be due to sea currents for sensors placed at sea, wind drift for sensors are dropped from planes or other errors due to placement from catapults or missiles (see e.g. Akyildiz et al. (2002)). Here we model the uncertainty in the locations as: $a_i(\zeta) = \bar{a}_i + \hat{a}_i\zeta_i$, where $\hat{a}_i \in \mathbb{R}_+$ is the maximal (absolute) deviation from the nominal value $\bar{a}_i$ for all $i \in L$. The uncertain parameter $\zeta = (\zeta_1, \dots, \zeta_{|L|})^\top$, where $\zeta_i \in \mathbb{R}^2$ for all $i \in L$, resides in a lifted budget uncertainty set $\mathcal{U}$ defined by

$$\mathcal{U} = \left\{(\zeta, \xi) : \ \zeta \le \xi, \ -\zeta \le \xi, \ \xi \le e, \ e^\top \xi \le \Gamma\right\},$$

where $\xi \in \mathbb{R}^{2|L|}$, $\|\cdot\|_1$ is the 1-norm and $\Gamma \ge 0$ is called the budget of uncertainty. Projecting $\mathcal{U}$ on the space of ζ, one can recover the classical budget uncertainty set $\{\zeta : \ -1 \le \zeta \le 1, \ \|\zeta\|_1 \le \Gamma\}$ of Bertsimas and Sim (2004). Some of the modules y_i need to be placed before the exact locations of $a_i(\zeta)$ are known, whereas others can be placed after. We define the set of indices H for those modules that have to be placed before the sensor locations are known. We associate a here-and-now variable $x_i \in \mathbb{R}^2$ for every $i \in H$. The two-stage model can now be written as:

$$\begin{aligned} \min_{x, y(\zeta,\xi)} \max_{(\zeta,\xi)\in\mathcal{U}} \quad & \sum_{(i,j)\in\mathcal{A}} \|y_i(\zeta,\xi) - y_j(\zeta,\xi)\|_2 \\ \text{s.t.} \quad & y_i(\zeta,\xi) = \bar{a}_i + \hat{a}_i\zeta_i \qquad \forall (\zeta,\xi) \in \mathcal{U} \ \forall i \in L \\ & y_i(\zeta,\xi) = x_i \qquad \forall (\zeta,\xi) \in \mathcal{U} \ \forall i \in H. \end{aligned} \tag{3.20}$$

The objective value of (3.20) gives the total energy required for the wireless transmissions in the network. One can eliminate the equality constraints, and then we have a static robust optimization model (i.e., without wait-and-see decisions). The static model is intractable in its current form because of convexity in the uncertain parameters in the objective. We propose to use consecutive dualization to derive equivalent linear reformulation of (3.20), and then solve the resulting model via linear decision rules.

3.5.2 Numerical setting

For the experiments we use two sets of data. For illustrative purposes we first consider a small instance with $N = 14$ points, of which 8 nominal sensor locations are uncertain and 6 modules need to be placed. For this, data from Boyd and Vandenberghe (2004, Section 8.7.3) is used, see Figure 3.2.[1] The maximal deviation from

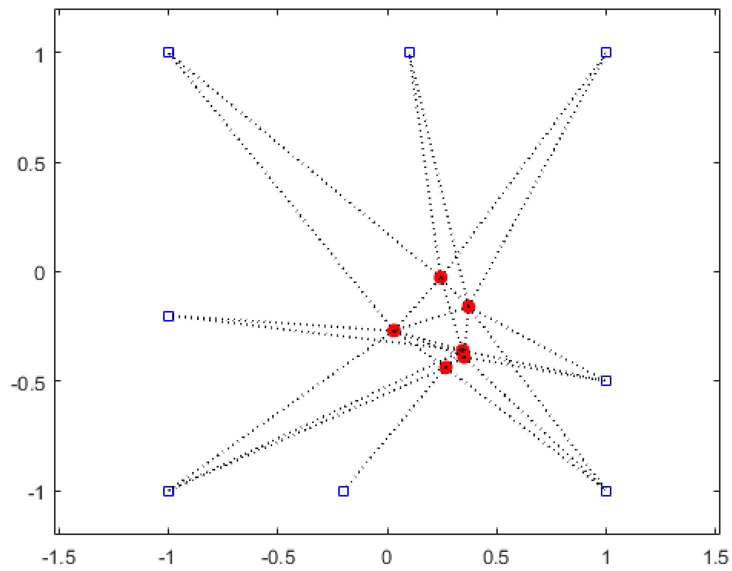


Figure 3.2 – Nominal solution from Boyd and Vandenberghe (2004, Figure 8.16).

the nominal locations is taken to be $\hat{a}_i \in \{0, 0.2, 0.5, 1\}$ for all $i \in L$. For instance for $\hat{a}_i = 0.2$, $i \in L$, Figure 3.3 illustrates the robust solution obtained from solving the dualized formulation of (3.20) with linear decision rules.

The second set of data shows results for larger instances. We choose $|L| \in \{10, ..., 70\}$ nominal sensor locations $\bar{a}_i$, $i \in L$ uniformly at random from $[-1, 1]^2$. The maximal deviation from the nominal locations is $\hat{a}_i = 0.3$ for all $i \in L$. We have $0.4|L|$ modules that have to collect data from the sensors. Each module is randomly linked with $|L|$ sensors. The modules are randomly linked into a cycle. We link the sensors that

[1]Data is obtained from the CVX website http://web.cvxr.com/cvx/examples.

are not connected with each of the $0.4|L|$ modules. We use the same computer and optimization software as mentioned in Section 3.4. Figure 3.4 depicts the robust solutions for $|L| = 30$.

Table 3.4 – Sensor network model with $N = 14$. P is obtained from solving (3.20). D is obtained from solving the dualized formulation of (3.20) with linear decision rules. LB-P-1 denotes the lower bounds obtained from one primal scenario. Time(s) reports the computation time (in seconds) for solving D.

$\hat{a}_i$	0	0.2	0.5	1
D	21.91	23.86	26.88	32.07
LB-P-1	21.91	23.79	26.64	31.46
Time(s)	0.1	0.1	0.1	0.1

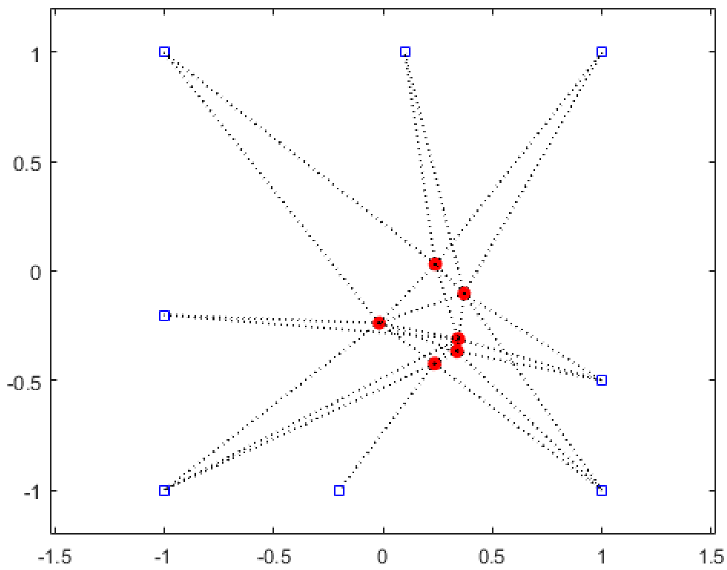


Figure 3.3 – Robust solution for $\hat{a}_i = 0.2$, $i \in L$, with the nominal sensor locations.

3.5.3 Results

For the small instance with $N = 14$ points, as $\hat{a}_i$, $i \in L = \{1, ..., 8\}$, increases, the obtained objective value from solving the dualized formulation of (3.20) with linear decision rules becomes larger (see Table 3.4), and the optimal module locations are more spread out (see Figure 3.3). Since the problem (3.20) has right-hand-side uncertainty, we focus on the lower bounds obtained from the primal scenarios. The lower bounds LB-P-1 are obtained from solving (3.11) with only one primal scenario. This primal scenario is obtained from the (dual) binding scenario of the first constraint of model (3.12). The differences between the lower bounds LB-P-1 and the upper bounds (Row D in Table 3.4) are within 1%. This indicates that the optimality gap of the obtained objective values is at most 1% of the optimal value. Here we do not consider Fourier-Motzkin elimination as it does not improve the obtained solutions within 1 hour. For medium and large instances considered in Table 3.5, the dualized formulation of (3.20) with linear decision rules can be computed efficiently, and the obtained solutions are near optimal.

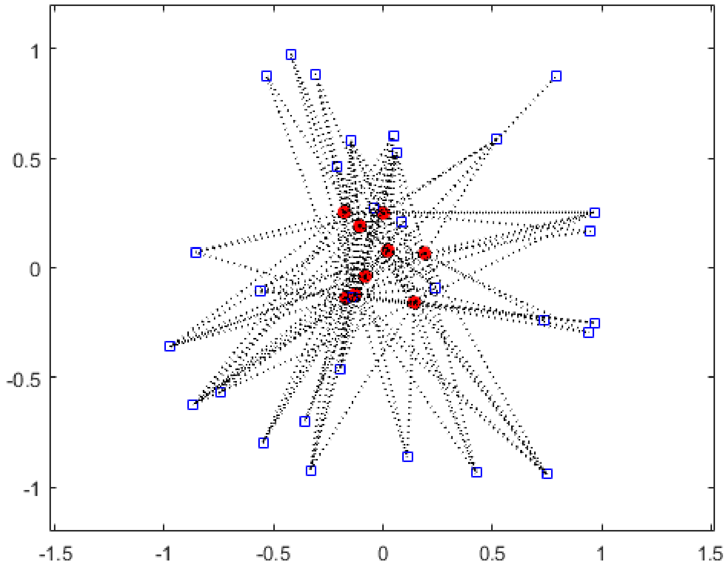


Figure 3.4 – Robust solution for $\hat{a}_i = 0.3$, $i \in L$, and $|L| = 30$, with the nominal sensor locations.

3.6 Conclusions and further research

In this chapter, we considered two-stage nonlinear robust optimization models with fixed recourse and polyhedral uncertainty set. We focus on models with convex wait-and-see decisions, for which a consecutive dual approach is developed to derive equivalent reformulations. Since the second stage variables appear linearly in the resulting model, many popular methods can be applied, e.g., linear decision rules and Fourier-Motzkin elimination. Moreover, we show that a broad class of static robust optimization models with convex uncertainties can be written into two-stage

Table 3.5 – Sensor network model with $\hat{a}_i = 0.3, i \in L$, where $|L| = \{10, ..., 70\}$. D is obtained from solving the dualized formulation of (3.20) with linear decision rules. LB-P-1 denotes the lower bounds obtained from one primal scenario. Time(s) reports the computation time (in seconds) for solving D. All the numbers are the average of 10 randomly generated instances.

$\lvert L \rvert$	10	20	30	40	50	60	70
D	18	50	104	171	260	381	520
LB-P-1	18	50	103	170	260	379	518
Time(s)	0.5	3	12	32	74	125	219

adjustable robust optimization models with convex recourse decisions, for which our dual approach can be applied. For the two numerical experiments, we use Fourier-Motzkin elimination and linear decision rules to find solutions for the proposed reformulations. A new lower bound method is introduced to provide empirical evidence that our approach gives near optimal solutions for the considered experiments.

On a theoretical level, one immediate future research direction would be to extend our approach to multistage robust models. In Chapter 2 we have shown the equivalence of the primal model and dualized model even after linear decision rules are imposed for both models. We would like to further investigate the dualized model with linear decision rules to find out the corresponding decision rules of the equivalent primal model.

On a numerical level, we would like to extend the finite adaptability approaches of Postek and den Hertog (2016) and Bertsimas and Dunning (2016) to solve two-stage nonlinear robust optimization models, and investigate the performance of finite adaptability approaches for the primal and its dualized formulations.

CHAPTER 4

Tractable nonlinear decision rules for robust optimization

4.1 Introduction

In many real world problems, optimization models contain parameters that are uncertain. Two approaches that deal with optimization problems that have uncertain parameters are stochastic and robust optimization. In stochastic optimization one requires probabilistic information on the uncertain parameter and optimizes expected objective values. We refer to the book by Kall and Wallace (1994) for more information on stochastic optimization. If the objective value is indeed an expectation, stochastic optimization is arguably the first choice one would make as paradigm. However, it is often computationally demanding and the required information on the probability distribution is not always available. The robust optimization paradigm uses uncertainty sets to model the uncertain parameters. These uncertainty sets are convex sets and allow the model to be reformulated into a tractable robust counterpart model, which is a normal convex optimization model, see Ben-Tal and Nemirovski (1998), Ben-Tal and Nemirovski (1999), Ben-Tal and Nemirovski (2002), El Ghaoui and Lebret (1997), El Ghaoui et al. (1998), Bertsimas et al. (2011b), Ben-Tal et al. (2015), and Bertsimas and Sim (2004). One of the key benefits of robust optimization models is that the tractable robust counterpart models can be solved efficiently with modern-day (commercial) solvers.

In robust optimization all decisions are being made *here-and-now* before the realization of the uncertain parameter is known. Sometimes there are also *wait-and-see* decisions that can be decided upon after (part of) the exact value of the uncertain parameter is known. In Ben-Tal et al. (2004), the robust optimization framework was extended to adjustable robust optimization that incorporates this type of wait-and-see decisions in the model. These wait-and-see decisions can be seen as functions, or decision rules, of the uncertain parameter: for each realization ζ a value has to be assigned to the wait-and-see decision. Unfortunately, optimization over the class of all decision rules is NP-hard as proven in Guslitzer (2002). Therefore, Ben-Tal et al. (2004) introduce the class of *linear decision rules* for adjustable robust optimization. These decision rules are affine functions of the uncertain parameter where coefficients of the linear decision rule are new here-and-now decisions. The resulting models with linear decision rules are again standard robust optimization models and can be reformulated into standard convex optimization models using robust optimization techniques such as Ben-Tal et al. (2015). Adjustable robust optimization and linear decision rules have been used in many applications such as: inventory management (Ben-Tal et al. 2005), facility location problems (Ardestani-Jaafari and Delage 2016b), network planning (Ng and Sy 2014) and lot-sizing on a network (Chapter 2)). More applications can be found in the survey papers by Bertsimas et al. (2011b) and Gabrel et al. (2014a).

Linear decision rules perform quite well in practice and are even the optimal decision rules in specific cases (Bertsimas et al. 2010; Iancu et al. 2013; Ardestani-Jaafari and Delage 2016b; Gounaris et al. 2013). However, linear decision rules are suboptimal in general and can even be a factor n off from the optimal objective value, where n is the number of uncertain parameters (Chen and Zhang 2009, Example 1). In Bertsimas and Goyal (2012) it is established that, under some structural model assumptions, linear decision rules give an $O(\sqrt{n})$ performance guarantee on the optimal objective value. This bound was slightly tightened for some particular uncertainty sets in Bertsimas and Bidkhori (2015), but remained $O(\sqrt{n})$ for many sets such as ellipsoidal uncertainty sets.

There have been a few papers that describe nonlinear decision rules, a richer class than linear decision rules. In Ben-Tal et al. (2009, Chapter 14) the authors explain how models with quadratic decision rules can be solved (approximately) by a semidefinite optimization problem. An even richer hierarchy of polynomial decision rules was introduced in Bertsimas et al. (2011a), which were also solved with semidefinite optimization models. For all these models, the tractable robust counterpart model when employing pure linear decision rules is a second-order cone optimization model. Therefore, enriching the decision rules to quadratic or higher order polynomials incurs additional complexity as the optimization class becomes semidefinite programming instead of second-order cone programming. A piecewise linear decision rule was introduced in Ben-Tal et al. (2016). They show that their piecewise linear decision rule outperforms, both theoretically and numerically, the pure linear decision rules. However, the results in that paper hold only for adjustable robust optimization models with a specific structure, such as right-hand side uncertainty and sign restrictions on the coefficients in the constraints.

Another approach relies on lifting the uncertainty set. After lifting the uncertainty set, the linear decision rule can incorporate the auxiliary parameters that are used to describe the lifted uncertainty set. For adjustable robust optimization these lifted uncertainty sets were first considered in Chen and Zhang (2009), where the new auxiliary parameters in their lifted uncertainty set represent the positive and negative part of the uncertain parameter. A similar lifting approach was considered in several stochastic or distributionally robust setting in Chen et al. (2008), See and Sim (2010), Goh and Sim (2010), Georghiou et al. (2015), and Bertsimas et al. (2017b).

In Section 4.3 we introduce nonlinear decision rules and show that they can be solved efficiently by tractable robust counterparts given in Proposition 4.2 using a convex hull description of a lifted uncertainty set. A closely related result to Proposition 4.1 that provides the convex hull, as well as the robust counterpart in Proposition 4.2, is given in Ben-Tal et al. (2011a). In that paper the result can also hold for uncertainty sets that are the intersections of uncertainty under some specific restrictions. The way that equivalence to a specific tractable robust optimization models is proven in this chapter is much simpler as it relies on the simple fact that the convex hull of the boundary of a convex set is the convex set itself. Furthermore, in Ben-Tal et al. (2011a) no indications on the performance, theoretical or numerical, of these nonlinear decision rules are given. We are also not aware of any other paper that uses, let alone test, the performance of those nonlinear decision rules. In this chapter we prove an a-priori theoretical bound on the performance of the nonlinear decision rules which are stronger than the equivalent bounds for pure linear decision rules known in the literature. Via two numerical examples, taken directly from the literature on adjustable robust optimization, we compare the performance of our nonlinear decision rules to pure linear decision rules and some other nonlinear decision rules from the literature. We also show how we can further lift the uncertainty set such that it strictly improves over the class of decision rules introduced in Chen and Zhang (2009). To summarize, the contributions of this chapter are the following:

1. We show how adjustable robust optimization models with nonlinear decision rules can be solved efficiently by solving an equivalent model with decision rules that are linear in the original uncertain parameter and in the auxiliary variables from a lifted uncertainty set. The major benefit of our approach is that the resulting models with the nonlinear decision rules can be solved exactly with a model that is of the same optimization class as the model with pure linear decision rules. We use a simpler proof for this than an equivalent result in Ben-Tal et al. (2011a).

2. We show that nonlinear decision rules for ellipsoidal and p-norm uncertainty sets are respectively an $O(n^{1/4})$ and $O(n^{(p-1)/p^2})$ approximation of the optimal objective value of the adjustable robust model. This improves upon the best known theoretical performance of $O(n^{1/p})$ for pure linear decision rules.

3. By further lifting the uncertainty set, we show how we can strictly improve

on the lifted decision rules introduced by Chen and Zhang (2009). We do this by showing that our class of nonlinear decision rules also includes the class of decision rules considered in that paper as a special case.

4. We demonstrate the power of the new nonlinear decision rules on two applications taken from the literature on adjustable robust optimization: random instances generated as in Ben-Tal et al. (2016) and the inventory model with flexible commitments from Ben-Tal et al. (2005). Via these examples, we show that our nonlinear decision rules strictly improve over pure linear decision rules, the decision rules from Chen and Zhang (2009) and the piecewise linear decision rules from Ben-Tal et al. (2016).

The rest of this chapter is organized as follows. In Section 4.2 we introduce the adjustable robust optimization model, linear decision rules and give some examples of convex uncertainty sets. Section 4.3 introduces our nonlinear decision rules and explains how these models can be efficiently solved. Our theoretical bounds for the nonlinear decision rules are given in Section 4.4. We show how to strictly improve on the decision rules from Chen and Zhang (2009) by further lifting the uncertainty set in Section 4.5. Our numerical results for the random instances and the inventory model with flexible commitments are presented in respectively Section 4.6 and 4.7.

Notation. We use Conv($\mathcal{U}$) to denote the convex hull of a set $\mathcal{U}$. Vectors and scalars are denoted by small letters and matrices by capital letters. We use subscript indices to indicate different elements of a vector or matrix. D_j refers to the j-th row of D and $D_j^\top$ to its j-th column. $D_{j,i}$ is the element in the j-th row and i-th column of D. The vector $\mathbf{1} \in \mathbb{R}^n$ denotes the vector of all ones of dimension n. The vector e_i denotes the vector in $\mathbb{R}^n$ with a one at the i-th entry and zeros elsewhere.

4.2 Adjustable robust optimization model

In an adjustable robust optimization model we must first make here-and-now decisions $x \in \mathbb{R}^{n_x}$ before we observe the realization of an uncertain parameter within a convex uncertainty set $\mathcal{U}$. After that, we can choose our wait-and-see decision $y \in \mathbb{R}^{n_y}$ to satisfy all the constraints. A general linear adjustable robust model has the form as in (1.6) introduced in Chapter 1:

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \forall \zeta \in \mathcal{U} \; \exists y : \; A(\zeta)x + By \geq D\zeta + d \\ & x \in \mathcal{X}, \end{aligned} \tag{4.1}$$

where $\mathcal{U} \subset \mathbb{R}^n$ is a compact convex uncertainty set. The uncertainty affects both the right-hand side via the matrix $D \in \mathbb{R}^{m \times n}$ and the left-hand side via $A(\zeta) : \mathbb{R}^n \to$

$\mathbb{R}^{n_x \times m}$, which is an affine function of ζ. The case where $A(\zeta)$ does not depend on ζ is called right-hand side uncertainty. We have $c \in \mathbb{R}^{n_x}$, $d \in \mathbb{R}^m$ and fixed recourse matrix $B \in \mathbb{R}^{m \times n_y}$. The set $\mathcal{X} \subset \mathbb{R}^{n_x}$ is a set describing additional restrictions on the here-and-now decisions x. For ease of exposition, we did not include any wait-and-see decisions in the objective. Nevertheless, model (4.1) can also be used to describe models with uncertainty or wait-and-see decisions in the objective by replacing the objective by an auxiliary variable $F \in \mathbb{R}$ and adding the constraint

$$c^\top x + b^\top y \leq F.$$

We can see model (4.1) as a model that optimizes over functions $y(\cdot)$, an infinite-dimensional space, because for every $\zeta \in \mathcal{U}$ we have to choose a different value of $y(\zeta)$. Fortunately, we can find good solutions for the model using linear decision rules as introduced in Ben-Tal et al. (2004). Instead of optimizing $y(\cdot)$ over all possible functions, linear decision rules restrict the class of functions to

$$y(\zeta) = \bar{y} + Y\zeta, \tag{4.2}$$

where $\bar{y} \in \mathbb{R}^{n_y}$ and $Y \in \mathbb{R}^{n_y \times n}$ is a vector and a matrix whose entries are new (here-and-now) decision variables that represent the coefficients of the linear decision rule. Substituting this linear decision rule in (4.1) we obtain

$$\begin{aligned} \min_{x,\bar{y},Y} \quad & c^\top x \\ \text{s.t.} \quad & \forall \zeta \in \mathcal{U}: \ A(\zeta)x + B(\bar{y} + Y\zeta) \geq D\zeta + d \\ & x \in \mathcal{X}, \end{aligned} \tag{4.3}$$

which is a standard robust optimization model without wait-and-see decisions similar to (1.8) in Chapter (1). Model (4.3) is referred to as the *Affine Adjustable Robust Counterpart* (AARC) and can be reformulated to a tractable convex optimization model for many choices of $\mathcal{U}$ using robust optimization techniques from Ben-Tal et al. (2015). For the remainder of this chapter we use the following description of the uncertainty set:

$$\mathcal{U} = \left\{ \zeta : \ Q\zeta \leq q, \ \sum_{i=1}^{n} g_i(\zeta_i) \leq \Gamma \right\}, \tag{4.4}$$

where $g_i : \mathbb{R} \to \mathbb{R}$ is a convex function, $Q \in \mathbb{R}^{n_p \times n}$ and $q \in \mathbb{R}^{n_p}$. The parameter Γ can be used to control the protection level and conservativeness of a solution. Higher values of Γ protect against more realizations of ζ, but could lead to more conservative solutions. Note that each function g_i only depends on one uncertain parameter ζ_i for all $i = 1, \ldots, n$. There are many uncertainty sets that are of this form, such as the ellipsoidal uncertainty set

$$\mathcal{U} = \left\{ \zeta : \ \sum_{i=1}^{n} \zeta_i^2 \leq \Gamma \right\}, \tag{4.5}$$

which is of the format (4.4), because we can take $g_i(\zeta_i) = \zeta_i^2$ for all $i = 1, \dots, n$. General ellipsoids of the form $\left\{\zeta : \ \zeta^\top Q \zeta \leq \Gamma\right\}$, with $Q \in \mathbb{R}^{n\times n}$ positive semidefinite, can be written in the format of (4.4) by including auxiliary variables $\xi \in \mathbb{R}^n$:

$$\mathcal{U} = \left\{(\zeta, \xi) : \ L\zeta \leq \xi, \ -L\zeta \leq \xi, \ \sum_{i=1}^{n} \xi_i^2 \leq \Gamma\right\},$$

where $L \in \mathbb{R}^{n\times n}$ is the Cholesky factor of Q, i.e., $L^\top L = Q$. There are many more uncertainty sets of the form (4.4) used in the literature such as p-norm uncertainty sets or the ϕ-divergence uncertainty sets from Ben-Tal et al. (2013)

$$\mathcal{U} = \left\{\zeta : Q\zeta \leq q, \ I_\phi(\zeta, \hat{\zeta}) \leq \Gamma\right\} \tag{4.6}$$

that can be used in a more distributionally robust setting for ambiguous discrete probability distribution sets. Here $I_\phi(\zeta, \hat{\zeta})$ is the ϕ-divergence that indicates the distance between a nominal distribution, given by the certain vector of nominal probabilities $\hat{\zeta}$, and the uncertain vector of probabilities ζ. For all ϕ-divergence examples given in Ben-Tal et al. (2013, Table 2) (Kullbeck-Leibner, Burg entropy, χ^2-distance, ...) the function $I_\phi(\zeta, \hat{\zeta})$ is of the form $\sum_{i=1}^n g_i(\zeta_i)$, so that (4.6) is of the form (4.4). A few other examples can be found in Bertsimas et al. (2017a), where data-driven uncertainty sets are formed, but there are also uncertainty sets in that paper that do not fit the format of (4.4).

4.3 Exact tractable models with nonlinear decision rules

Throughout this section we consider the case with $Q = 0$ and $q = 0$ in (4.4). Hence, the uncertainty set has the form

$$\mathcal{U} = \left\{\zeta : \ \sum_{i=1}^{n} g_i(\zeta_i) \leq \Gamma\right\}. \tag{4.7}$$

Instead of solving the problem with linear decision rules, we use the following decision rule:

$$y(\zeta, u) = \bar{y} + Y\zeta + Wu, \qquad \text{where } \zeta \in \mathcal{U} \text{ and } u_i = g_i(\zeta_i), \ i = 1, \dots, n. \tag{4.8}$$

Besides the variables $\bar{y} \in \mathbb{R}^{n_y}$ and $Y \in \mathbb{R}^{n_y\times n}$ we had in (4.2), we also have $W \in \mathbb{R}^{n_y\times n}$ as new additional variables for the coefficients of $g_i(\zeta_i)$ in the decision rule. Note that this decision rule is nonlinear when projected on ζ because of the terms $u_i = g_i(\zeta_i)$ for all $i = 1, \dots, n$. If we plug this nonlinear decision rule into the adjustable robust optimization model, then we obtain the following robust optimization model:

$$\begin{aligned} \min_{x,\bar{y},Y,W} \quad & c^\top x \\ \text{s.t.} \quad & \forall(\zeta, u) \in \mathcal{V}^{ext} : \ A(\zeta)x + B\left(\bar{y} + Y\zeta + Wu\right) \geq D\zeta + d \\ & x \in \mathcal{X}, \end{aligned} \tag{4.9}$$

where we have used a new extended uncertainty set

$$\mathcal{V}^{ext} = \left\{ (\zeta, u) : \; g_i(\zeta_i) = u_i, \; i = 1, \ldots, n, \; \sum_{i=1}^{n} u_i \leq \Gamma \right\}. \tag{4.10}$$

This almost seems to be a tractable robust optimization problem, but the issue is that the extended uncertainty set is not convex. Fortunately, since (4.9) is a linear robust optimization model, the feasible region does not change if we replace the uncertainty set $\mathcal{V}^{ext}$ by its convex hull. In general, it is hard to determine the convex hull of a nonconvex set and exponentially many constraints may be needed for its description. For the uncertainty sets of the form (4.7) the convex hull allows for a small description without any additional parameters as stated in the next proposition.

Proposition 4.1 *Given that $\mathcal{U}$ is compact, we have*

$$Conv(\mathcal{V}^{ext}) = \mathcal{V},$$

where

$$\mathcal{V} = \left\{ (\zeta, u) : \; g_i(\zeta_i) \leq u_i, \; i = 1, \ldots, n, \; \sum_{i=1}^{n} u_i \leq \Gamma \right\}. \tag{4.11}$$

Proof. The only difference between $\mathcal{V}^{ext}$ and $\mathcal{V}$ is that in the former set $g_i(\zeta_i) \leq u_i$ holds with equality for all $i = 1, \ldots, n$. We have $\mathcal{V}^{ext} \subset \mathcal{V}$ and $\mathcal{V}$ bounded because $\mathcal{U}$ is compact, so all we have to prove is that for each extreme point of $\mathcal{V}$ we have $g_i(\zeta_i) = u_i$ for all $i = 1, \ldots, n$. Suppose, for the sake of contradiction, that this is not the case, so there exists an extreme point and $i \in \{1, \ldots, n\}$ such that $g_i(\zeta_i) < u_i$. Since the functions g_i are real valued and convex, they are also continuous for all $i = 1, \ldots, n$. Hence, we can take an $\epsilon > 0$ such that $g_i(\zeta_i + \epsilon) \leq u_i$ and $g_i(\zeta_i - \epsilon) \leq u_i$. The points $(\zeta + \epsilon e_i, u)$ and $(\zeta - \epsilon e_i, u)$ are distinct elements of $\mathcal{V}$ and

$$(\zeta, u) = \tfrac{1}{2}(\zeta - \epsilon e_i, u) + \tfrac{1}{2}(\zeta + \epsilon e_i, u).$$

Hence, the point (ζ, u) can be written as the convex combination of two distinct points in $\mathcal{V}$. Therefore, (ζ, u) is not an extreme point which leads to the desired contradiction. ■

In this chapter we call $\mathcal{V}$ in (4.11) the lifted uncertainty set. An example of a lifted uncertainty sets is

$$\mathcal{V} = \left\{ (\zeta, u) : \; |\zeta_i|^p \leq u_i, \; i = 1, \ldots, n, \; \sum_{i=1}^{n} u_i \leq \Gamma \right\}, \tag{4.12}$$

which is just an instance of (4.11) that lifts the p-norm uncertainty set. Below we provide the tractable robust counterpart for the ellipsoidal uncertainty set ($p = 2$) in

(4.12), which is used in the numerical examples in this chapter. For ease of exposition we present here the model for the case where $A(\zeta)$ does not depend on ζ (right-hand side uncertainty). For the general case where $A(\zeta)$ is a linear function of ζ, as well as for general functions g_i in the uncertainty set (4.7), we refer to Appendix 4.A.

Proposition 4.2 *$(x, \bar{y}, Y, W)$ is feasible (and optimal) for* (4.9) *with right-hand side uncertainty and the lifted ellipsoidal ($p = 2$) uncertainty set* (4.12) *if and only if there exist $\lambda \in \mathbb{R}^{n\times m}$, $\psi \in \mathbb{R}^m$ and $t \in \mathbb{R}^{n\times m}$ such that $(x, \bar{y}, Y, W, \psi, \lambda, t)$ is feasible (and optimal) for*

$$
\begin{aligned}
\min_{x,\bar{y},Y,W,\psi,\lambda,t} \quad & c^\top x \\
\text{s.t.} \quad & A_j x + B_j \bar{y} - \sum_{i=1}^{n} t_{i,j} - \Gamma \psi_j \geq d_j \qquad j = 1, \ldots, m \\
& \left\| \left(2\left(Y_i^\top B_j - D_{j,i} \right), 4\lambda_{i,j} - t_{i,j} \right) \right\|_2 \leq 4\lambda_{i,j} + t_{i,j} \qquad \forall i = 1, \ldots, n, \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = 1, \ldots, m \\
& - W_i^\top B_j + \lambda_{i,j} \leq \psi_j \qquad \forall i = 1, \ldots, n,\ j = 1, \ldots, m \\
& \lambda_{i,j} \geq 0 \qquad \forall i = 1, \ldots, n,\ j = 1, \ldots, m \\
& \psi_j \geq 0 \qquad \forall j = 1, \ldots, m \\
& x \in \mathcal{X}.
\end{aligned}
\tag{4.13}
$$

Proof. Since model (4.9) is a linear robust optimization model we can replace $\mathcal{V}^{ext}$ by its convex hull $\mathcal{V}$ given in (4.11). Then the robust model (4.9) has a convex uncertainty set and the robust counterpart can be derived using standard robust methods as described in Ben-Tal et al. (2015). For completeness we give the full derivation of the tractable robust counterpart in Appendix 4.A. ■

Model (4.13) is a second-order cone model because of the Euclidean norm in the constraints. The tractable robust counterpart formulation of (4.3), using pure linear decision rules, is also a second-order cone model, see for details and full derivation Ben-Tal et al. (2009). The difference is that model (4.13) has mn second-order cone constraints with a norm of a two-dimensional vector, whereas the version with linear decision rules would only have m second-order cone constraints with norms of a n-dimensional vector. Of course, the nonlinear decision rule also adds more variables, so (4.13) takes most likely more computational effort to solve, but the optimization class is still the same. However, we argue that the gain in objective value is often worth the small additional computational effort as shown in our numerical examples in Sections 4.6 and 4.7. The result in Proposition 4.2 has been proven before by Ben-Tal and den Hertog (2014, p27-28). They use a different proof that uses more advanced convex analysis, whereas we only have to rely on the more simple facts that the convex hull of the boundary of a compact convex set is the convex set itself. Their result holds in some additional cases as well, given that they can ensure that some extra structural restrictions are satisfied.

Linear decision rules of the form (4.2) can be extended from two-stage models to multi-stage models. In multi-stage models one has to take nonanticipativity of the decision rules into account. Nonanticipativity means that a decision cannot rely on information that is only revealed in later stages. Pure linear decision rules can be made nonanticipative by forcing some of the entries of V in (4.2) to zero. For instance, if the k-th decision $y_k(\zeta)$ is only allowed to depend on all information from the first $k-1$ elements in ζ, then this is achieved by requiring all variables in the upper triangular part of V (including the diagonal) to be zero. Similarly, for our nonlinear decision rules we additionally require that the upper triangular part of W in (4.8) is zero. This can be done since the auxiliary parameters u_i are determined only by the value of $g_i(\zeta_i)$ for all $i = 1, \ldots, n$, which is known whenever ζ_i is revealed.

Remark. The exactness of the convex hull in Theorem 4.1 only holds for the case where $Q = 0$ or $q = 0$. Otherwise, the set $\mathcal{V}$ will be larger than the convex hull of the extreme points. For instance, if we consider the ellipsoidal set (4.5) intersected with the nonnegative orthant, then the set $\mathcal{V}^{ext}$ would be given by

$$\mathcal{V}^{ext} = \left\{ (\zeta, u) : \ \zeta_i \geq 0, \ \zeta_i^2 = u_i, \ i = 1, \ldots, n, \ \sum_{i=1}^{n} u_i \leq \Gamma \right\}. \tag{4.14}$$

The description of the convex hull of this set is much more difficult since it is restricted to the nonnegative orthant. The convex hull of the set $\mathcal{V}^{ext}$ in (4.14) is contained in the set

$$\mathcal{V} = \left\{ (\zeta, u) : \ \zeta_i \geq 0, \ \zeta_i^2 \leq u_i, \ i = 1, \ldots, n, \ \sum_{i=1}^{n} u_i \leq \Gamma \right\}, \tag{4.15}$$

which also has extreme points with $\zeta = 0$ and $u = e_i$, $i = 1, \ldots, n$. These points do not correspond to $u_i = \zeta_i^2$. Therefore, the set $\mathcal{V}$ in (4.15) is larger than the convex hull of $\mathcal{V}^{ext}$. This lifted uncertainty set can still be used to solve the model with nonlinear decision rules of the form (4.8). However, since $\mathcal{V}$ is larger than the convex hull, the solutions might be conservative and not result in the optimal decision rules of the form (4.8).

4.4 Theoretical bound for ellipsoidal and p-norm sets

For ellipsoidal and general p-norm uncertainty sets we provide a bound on the performance of the nonlinear decision rules. Let F_{ARO} be the optimal value of (4.1) and let F_{NDR} be the optimal value of (4.9), the model with nonlinear decision rules. In this section we restrict ourselves to right-hand side uncertainty, that is $A(\zeta) = A$ for some $A \in \mathbb{R}^{m \times n_x}$, meaning that the left-hand side of the constraints does not depend on ζ.

Theorem 4.1 *Let $\mathcal{X}$ be a convex cone, $A(\zeta) = A$ for some $A \in \mathbb{R}^{m \times n_x}$, $D \in \mathbb{R}_+^{m \times n}$, $d \in \mathbb{R}_+^m$ and for $p \geq 2$, $\mathcal{U}$ the p-norm uncertainty set (possibly intersected with the nonnegative orthant) with lifted uncertainty set* (4.12)*. Then we have*

$$F_{ARO} \geq \tfrac{2}{p}(p-1)^{(p-1)/p} n^{(p-1)/p^2} F_{NDR}.$$

Proof. Without loss of generality we assume that $\Gamma = 1$ in (4.12). If $\Gamma \neq 1$, then we can always scale D with Γ and take the budget parameter equal to 1. We prove the theorem initially for the case where the p-norm uncertainty set $\mathcal{V}$ is as in (4.12). First we consider a slightly adapted version of the model (4.1):

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & \forall (\zeta, u) \in \mathcal{V} \ \exists y : \ Ax + By \geq Du + d \\ & x \in \mathcal{X}. \end{aligned} \tag{4.16}$$

This model is equivalent to (4.1), but with the uncertain parameter ζ replaced by u in the constraints. Since ζ does not appear in (4.16), we can also replace $\mathcal{V}$ by its projection on u which is the simplex set $\Lambda = \{u \geq 0 : \sum_{i=1}^n u_i \leq 1\}$. The unit simplex set is contained in the $p-$norm set $\mathcal{U}$ so the optimal objective value of (4.16) is a lower bound for the optimal objective value of (4.1). For simplex sets, linear decision rules are optimal (Bertsimas and Goyal 2012), so there exists a solution $(\hat{x}, \hat{y}(u) = \bar{y} + \bar{W}u)$ that gives the optimal objective value for (4.16). Now consider the *nominal* model of (4.1) which we define as

$$\begin{aligned} \min_{x,y} \quad & c^\top x \\ \text{s.t.} \quad & Ax + By \geq D\tfrac{1}{n^{1/p}}\mathbf{1} + d, \\ & x \in \mathcal{X}, \end{aligned} \tag{4.17}$$

which is equal to (4.1), but with ζ replaced by $\frac{1}{n^{1/p}}\mathbf{1}$. Let F_{nom} be the optimal value of the nominal model. Since $\frac{1}{n^{1/p}}\mathbf{1} \in \mathcal{U}$ (just one scenario out of $\mathcal{U}$), we have that the optimal objective value of the nominal model (4.17) also gives a lower bound to the objective value of (4.1), i.e., $F_{nom} \leq F_{ARO}$. Let (x^{nom}, y^{nom}) denote the optimal solution of the nominal model (4.17). Define $\alpha = \frac{1}{p}(p-1)^{(p-1)/p} n^{(p-1)/p^2}$. We consider the following candidate solution $(x^*, y^*(u))$ for (4.9): $x^* = \alpha\left(\hat{x} + x^{nom}\right)$ and $y^*(u) = \alpha\left(y^{nom} + \hat{y}(u)\right) = \alpha\left(y^{nom} + \bar{y} + \bar{W}^* u\right)$. What remains to prove are the following two properties of the candidate solution:

i. $(x^*, y^*(u))$ is feasible for (4.9).

ii. The objective value of $(x^*, y^*(u))$ is at most 2α times larger than F_{ARO}.

(i. Feasibility).
For the candidate solution $(x^*, y^*(u))$ we have

$$\begin{aligned} Ax^* + By^*(u) &= \alpha\left(A\left(\hat{x} + x^{nom}\right)\right) + B\left(y^{nom} + \hat{y}(u)\right)) \\ &\geq \alpha\left(A\hat{x} + B\hat{y}(u) + D\tfrac{1}{n^{1/p}}\mathbf{1} + d\right) \\ &\geq \alpha\left(Du + d + D\tfrac{1}{n^{1/p}}\mathbf{1} + d\right) \\ &= D\alpha\left(u + \tfrac{1}{n^{1/p}}\mathbf{1}\right) + 2\alpha d, \end{aligned}$$

where the first inequality follows from the fact that (x^{nom}, y^{nom}) is feasible for (4.17) and the second inequality holds because $\hat{x}, \hat{y}(u)$ is feasible for (4.16). We note that for all $(\zeta, u) \in \mathcal{U}$ we have $u_i \geq |\zeta_i|^p$. Furthermore, we can show that

$$\alpha\left(|\zeta_i|^p + \tfrac{1}{n^{1/p}}\right) - \zeta_i \geq 0 \tag{4.18}$$

for all $\zeta_i \in \mathbb{R}$. The validity of the inequality in (4.18) can be shown by minimizing the left-hand side over $\zeta_i \in \mathbb{R}$. First note that for $\zeta_i \leq 0$ the inequality holds trivially, hence we only have to prove it for $\zeta_i > 0$. The first order conditions (with $\zeta_i > 0$) give the necessary and sufficient conditions for the minimizer $\bar{\zeta}_i$ of the function on the left-hand side of (4.18):

$$\alpha p \bar{\zeta}_i^{p-1} - 1 = 0 \qquad \Leftrightarrow \qquad \bar{\zeta}_i = \tfrac{1}{(\alpha p)^{1/(p-1)}} = \tfrac{1}{(p-1)^{1/p} n^{1/p^2}}.$$

Plugging the solution for $\bar{\zeta}_i$ from the first order conditions into (4.18), we get

$$\begin{aligned} \alpha\left(|\bar{\zeta}_i|^p + \tfrac{1}{n^{1/p}}\right) - \bar{\zeta}_i &= \tfrac{1}{p}(p-1)^{(p-1)/p} n^{(p-1)/p^2}\left(\tfrac{1}{(p-1)}\tfrac{1}{n^{1/p}} + \tfrac{1}{n^{1/p}}\right) - \tfrac{1}{(p-1)^{1/p} n^{1/p^2}} \\ &= 0. \end{aligned}$$

Hence, combining (4.18) with the fact that $|\zeta_i|^p \leq u_i$, we obtain for every $i = 1, \ldots, n$:

$$\alpha\left(u_i + \tfrac{1}{n^{1/p}}\right) \geq \zeta_i.$$

Therefore, since all entries of D and d are positive, we have

$$\begin{aligned} Ax^* + By^*(u) &\geq D\alpha\left(u + \tfrac{1}{n^{1/p}}\mathbf{1}\right) + 2\alpha d \\ &\geq D\zeta + 2\alpha d \\ &\geq D\zeta + d, \end{aligned}$$

where the last inequality follows from the fact that $\alpha \geq \frac{1}{2}$ for every $p \geq 2$ and integer n. Furthermore, since the set $\mathcal{X}$ is a convex cone and $\hat{x}, x^{nom} \in \mathcal{X}$, we have that $x^* = \alpha\hat{x} + \alpha x^{nom} \in \mathcal{X}$. Hence, the solution $(x^*, y^*(u))$ is feasible for (4.9).

(ii. Approximation factor).
The objective value of the solution $(x^*, y^*(u))$ is given by $c^\top x^*$. Since that is just a feasible solution, it is an upper bound of the optimal objective value with nonlinear decision rules F_{NDR}. Hence, we obtain

$$F_{NDR} \leq c^\top x^* = \alpha c^\top \left(\hat{x} + x^{nom}\right) \leq 2\alpha F_{ARO} = \tfrac{2}{p}(p-1)^{(p-1)/p^2} n^{(p-1)/p^2} F_{ARO}.$$

Finally, note that the entire proof holds verbatim if we intersect the p-norm uncertainty set with the nonnegative orthant. ■

The following corollary is an immediate implication for ellipsoidal uncertainty sets.

Corollary 4.1 *For ellipsoidal uncertainty sets ($p = 2$), $\mathcal{X}$ a conic set, $A(\zeta) = A$ for some $A \in \mathbb{R}^{m \times n_x}$ $D \in \mathbb{R}_+^{m \times n}$ and $d \in \mathbb{R}_+^m$, the optimal objective value of* (4.9) *gives an $O(n^{(1/4)})$ approximation of* (4.1).

For p-norms, the bound on the optimal objective value in Theorem 4.1 of $O(n^{(p-1)/p^2})$ for our nonlinear decision rules improves on the best known bound of $O(n^{1/p})$ for the performance of linear decision rules given in Bertsimas and Bidkhori (2015) and Bertsimas and Goyal (2012).

The bound in Theorem 4.1 depends on the dimension of the uncertain parameter, which is often far less than the number of constraints and variables. For $n \leq 16$, we see that it guarantees a 2-approximation for model (4.1) for ellipsoidal uncertainty sets, regardless of the dimensions of the wait-and-see decisions or the number of constraints.

The bound in Theorem 4.1 equals the bound given in Ben-Tal et al. (2016) of $O(n^{(p-1)/p^2})$. In that paper they give a specific piecewise linear decision rule construction. The models they consider are of slightly more restrictive setting than (4.1) as described above, since D is the identity matrix and $d = 0$ in that paper. We compare the numerical performance of their piecewise linear decision rule and our nonlinear decision rule (4.8) by using the same example as in Ben-Tal et al. (2016) in Section 4.6.

4.5 Further improvements for ellipsoidal and p-norms sets

The ellipsoidal ($p = 2$) and general p-norm uncertainty sets allow for higher dimensional liftings:

$$\mathcal{V}_p^{ext} = \left\{ (\zeta, u, v) : \; u_i = v_i^p, \; |\zeta_i| = v_i, i = 1, \ldots, n, \; \sum_{i=1}^n u_i \leq \Gamma \right\},$$

which then allows us to consider decision rules of the form

$$y(\zeta, u) = \bar{y} + Y\zeta + Wu + Zv, \qquad \text{where } u_i = v_i^p,\ v_i = |\zeta_i|,\ i = 1, \ldots, n \tag{4.19}$$

with $Z \in \mathbb{R}^{n_y \times n}$ additional here-and-now variables for the coefficients of v in the decision rule. This uncertainty set contains, next to the p-th power of the uncertain parameter $|\zeta_i|^p$, also the absolute values $|\zeta_i|$, for all $i = 1, \ldots, n$. To derive tractable robust counterparts, we need again a description of the convex hull of $\mathcal{V}_p^{ext}$. Unfortunately, we cannot give an efficient exact description, but the convex hull of $\mathcal{V}_p^{ext}$ is contained in the set

$$\mathcal{V}_p = \left\{ (\zeta, u, v) :\ v_i^p \leq u_i,\ |\zeta_i| \leq v_i, i = 1, \ldots, n,\ \sum_{i=1}^{n} u_i \leq \Gamma \right\}. \tag{4.20}$$

Note that $\mathcal{V}_p$ is not the exact convex hull of $\mathcal{V}_p^{ext}$ since it also has extreme points $(0, 0, e_i)$, $i = 1, \ldots, n$. Clearly, these additional extreme points do not correspond with $u_i = v_i^p$ for all $i = 1, \ldots, n$. However, we can show that for all extreme points of $\mathcal{V}_p$ we still have $v_i = |\zeta_i|$.

Property 4.1 *For each extreme points of $\mathcal{V}_p$ in (4.20) we have $v_i = |\zeta_i|$ for all $i = 1, \ldots, n$.*

Proof. Suppose, for the sake of contradiction, that there exists an extreme point (ζ, u, v) with $v_i > |\zeta_i|$ for some $i = 1, \ldots, n$. Then we can define $\epsilon_i = v_i - |\zeta_i|$ for all $i = 1, \ldots, n$. The extreme point is then equal to the convex combination $\frac{1}{2}(\zeta + \epsilon, u, v) + \frac{1}{2}(\zeta - \epsilon, u, v)$, where the two points are distinct and in $\mathcal{V}_p$, which leads to the desired contradiction. ■

Projecting $\mathcal{V}_p^{ext}$ onto (ζ, u) gives us again the lifted uncertainty set (4.12), for which the extreme points all satisfy $u_i = |\zeta_i|^p$ for all $i = 1, \ldots, n$ by Theorem 4.1. On the other hand, the projection on (ζ, v) gives us the following set:

$$\mathcal{W}_p = \left\{ (\zeta, v) :\ |\zeta_i| \leq v_i\ i = 1, \ldots, n\ \sum_{i=1}^{n} v_i^p \leq \Gamma \right\}.$$

We show that this projection is in some sense equivalent to the lifted uncertainty set proposed by Chen and Zhang (2009), which was defined as:

$$\widetilde{\mathcal{W}}_p = \left\{ (\zeta, v^-, v^+) :\ \zeta = v^+ - v^-,\ v^+ \geq 0,\ v^- \geq 0,\ \left\| v^+ + v^- \right\|_p \leq \Gamma^{1/p} \right\}$$

and used linear decision rules of the form

$$y(\zeta, v^+, v^-) = \bar{y} + Y\zeta + Z^+ v^+ + Z^- v^-, \tag{4.21}$$

where $Z^+, Z^- \in \mathbb{R}^{n_y \times n}$ are additional here-and-now variables for the coefficients of respectively v^+ and v^-. It can be shown that the lifted uncertainty sets $\mathcal{W}_p$ and $\widetilde{\mathcal{W}}_p$ are equivalent under a linear transformation.

Property 4.2

$$\widetilde{\mathcal{W}}_p = \left\{(\zeta, v^+, v^-) : \ v^+ = \tfrac{1}{2}(v+\zeta), \ v^- = \tfrac{1}{2}(v-\zeta), \ (\zeta, v) \in \mathcal{W}_p\right\}.$$

Proof. "$\supset$" Let $(\zeta, v) \in \mathcal{W}_p$, $v^+ = \frac{1}{2}(v+\zeta)$ and $v^- = \frac{1}{2}(v-\zeta)$. Then we directly have $\zeta = v^+ - v^-$ and $v^+, v^- \geq 0$. Furthermore, $\|v^+ + v^-\| = \|v\| \leq \Gamma^{1/p}$. Hence $(\zeta, v^+, v^-) \in \mathcal{W}_p$.

"$\subset$" Let $(\zeta, v^+, v^-) \in \mathcal{W}_p$. Define $v = 2v^+ - \zeta$. By the relation $\zeta = v^+ - v^-$ we also have $v = 2v^+ - \zeta = 2(\zeta + v^-) - \zeta = 2v^- + \zeta$. In other words $v^+ = \frac{1}{2}(v+\zeta)$ and $v^- = \frac{1}{2}(v-\zeta)$. Also, we have

$$v = 2v^+ - \zeta \geq \zeta,$$

as $v^+ = \zeta + v^- \geq \zeta$ and $v^- \geq 0$. Similary $v = 2v^- - \zeta \geq -\zeta$. Therefore, $v_i \geq |\zeta_i|$ for all $i = 1, \ldots, n$. Finally, $\|v\| = \|2v^+ - \zeta\| = \|2v^+ - (v^+ - v^-)\| = \|v^+ + v^-\| \leq \Gamma^{1/p}$, so $\sum_{i=1}^n v_i^p \leq \Gamma$. Hence, $(\zeta, v) \in \mathcal{W}_p$. ■

Given this equivalence under linear transformation, we can show that decision rules of the form (4.19) perform better than (4.21). To do so, we first show that the performance of linear decision rules cannot be improved when adding auxiliary variables that are direct affine transformations of the uncertain parameters.

Property 4.3 *Let $\mathcal{U}$ be any convex set and $\widehat{\mathcal{U}}$ be a lifted uncertainty set where the additional variables $\hat{v} \in \mathbb{R}^{\hat{n}}$ are direct affine transformations of ζ, i.e.,*

$$\widehat{\mathcal{U}} = \{(\zeta, \hat{v}) : \ \hat{v} = \omega + \Omega\zeta, \ \zeta \in \mathcal{U}\}$$

for some $\omega \in \mathbb{R}^{\hat{n}}$ and $\Omega \in \mathbb{R}^{\hat{n}\times n}$. Then for any decision rule $\hat{y}(\zeta, \hat{v})$ that is linear in ζ and $\hat{v}$, we can construct a decision rule $\tilde{y}(\zeta)$ that is linear in ζ with $\hat{y}(\zeta, \hat{v}) = \tilde{y}(\zeta)$ for all $\zeta \in \mathcal{U}$.

Proof. Consider a decision rule $\hat{y}(\zeta, \hat{v}) = \hat{z} + \hat{Y}\zeta + \hat{Z}\hat{v}$ for all $(\zeta, \hat{v}) \in \widehat{\mathcal{U}}$. Then we can construct a decision rule linear in ζ as $\tilde{y}(\zeta) = \tilde{z} + \tilde{Y}\zeta$, where $\tilde{z} = \hat{z} + \hat{Z}\omega$ and $\tilde{Y} = \hat{Y} + \hat{Z}\Omega$. For this decision rules we then have

$$\tilde{y}(\zeta) = \tilde{z} + \tilde{Y}\zeta = \left(\hat{z} + \hat{Z}\omega\right) + \left(\hat{Y} + \hat{Z}\Omega\right)\zeta = \hat{z} + \hat{Y}\zeta + \hat{Z}\hat{v} = \hat{y}(\zeta, \hat{v})$$

for all $\zeta \in \mathcal{U}$. ■

By Property 4.2 we know that the set $\widetilde{\mathcal{W}}_p$ only contains auxiliary variables that are an affine transformation of the parameters in $\mathcal{W}_p$. Hence, linear decision rules that use the auxiliary variables in $\widetilde{\mathcal{W}}_p$ cannot perform better than $\mathcal{W}_p$. Furthermore, since $\mathcal{W}_p$ is a projection of the set $\mathcal{V}_p$, the linear decision rules (4.19) that use all parameters in $\mathcal{V}_p$ (also the parameters u_i that are nonlinearly related) constitute a richer class of decision rules. Therefore, linear decision rules (4.19) based on the lifted uncertainty set $\mathcal{V}_p$ must perform at least as good as decision rules of the form (4.21) from Chen and Zhang (2009). In the numerical results in Section 4.7 we show that these decision rules strictly improve over decision rules of the form (4.21).

4.6 Example 1: Random instances

4.6.1 The model

We consider the same model as in Ben-Tal et al. (2016):

$$\begin{aligned} \min_{x,F} \quad & F \\ \text{s.t.} \quad & \forall \zeta \in \mathcal{U}\ \exists y : \begin{cases} F - c^\top x - d^\top y \geq 0 \\ Ax + By \geq \zeta \\ y \geq 0 \end{cases} \\ & x \geq 0, \end{aligned} \tag{4.22}$$

where $\mathcal{U} = \{\zeta \in \mathbb{R}^n_+ : \ \sum_{i=1}^n \zeta_i^2 \leq 1\}$, the ellipsoidal uncertainty set intersected with the nonnegative orthant. The same input as in Ben-Tal et al. (2016) is used:

$$A = B = I + G,$$

where I is the $n \times n$ identity matrix and $G_{ij} = |Y_{ij}|/\sqrt{n}$, Y_{ij} independently drawn from a standard normal distribution for all $i, j = 1, \ldots, n$. The objective coefficients are $c = d = \mathbf{1}$, the vector of all ones.

We first note that the decision rules with absolute values from Chen and Zhang (2009) cannot be applied to this model since the uncertainty set is restricted to the nonnegative orthant, so $|\zeta_i| = \zeta_i \geq 0$ for all $i = 1, \ldots, n$. Hence, the decision rules from Chen and Zhang (2009) would coincide with a pure linear decision rule. We solve the model using pure linear decision rules (4.2), the nonlinear decision rule (4.8) and the piecewise linear decision rules described in Ben-Tal et al. (2016). In that paper they replace the uncertainty set by a "dominating set", which can be used for two-stage models that have certain structure (such as the matrix A being nonnegative). The set that replaces the uncertainty set is given by

$$\bar{\mathcal{W}} = n^{1/4} \left\{ e_1, \ldots, e_n, \tfrac{1}{\sqrt{n}} \mathbf{1} \right\}.$$

Ben-Tal et al. (2016) showed that when this specific discrete set is used, they only have to solve a small linear program. They also show it is equivalent to a particular piecewise linear decision rule and that it gives an $O(n^{1/4})$ approximation guarantee.

Unfortunately, however, it is not guaranteed to do better than linear decision rules. Recall that the nonlinear decision rule (4.8) also has a theoretical approximation factor of $O(n^{1/4})$ for the optimal objective value of (4.22), whereas pure linear decision rules have an approximation factor of $O(\sqrt{n})$ (Bertsimas and Goyal 2012).

4.6.2 Numerical results

We use the algebraic modelling language JuMP created by Dunning et al. (2017) in the Julia programming language. We performed the experiments on an Intel i7-4770 3.40GHz Windows PC with 8GB of RAM. All trials use Mosek 8 as a solver. We define the excess as the percentage by which the objective value has decreased compared to the optimal objective value with linear decision rules:

$$\text{Excess('method')} = \left(\frac{\text{objective value 'method'}}{\text{objective value LDR}} - 1 \right) \cdot 100\%, \tag{4.23}$$

where 'method' can refer to any decision rule, or other method used to find solutions to the adjustable robust optimization model. Since we minimize costs, lower values of excess indicate better performance. Note that the Excess is sensitive to translations. This means that, if the objective function is shifted with a constant, one could make this value arbitrarily small or large. Throughout this paper we take the instances directly from the literature without applying any translations or other modifications that could enlarge or shrink the excess. As in Ben-Tal et al. (2016), we have used instance sizes $n = 10, 20, \ldots, 100$ and averaged the results over 100 random instances for each size. The results are given in Table 4.1 and are graphically depicted in Figure 4.1.

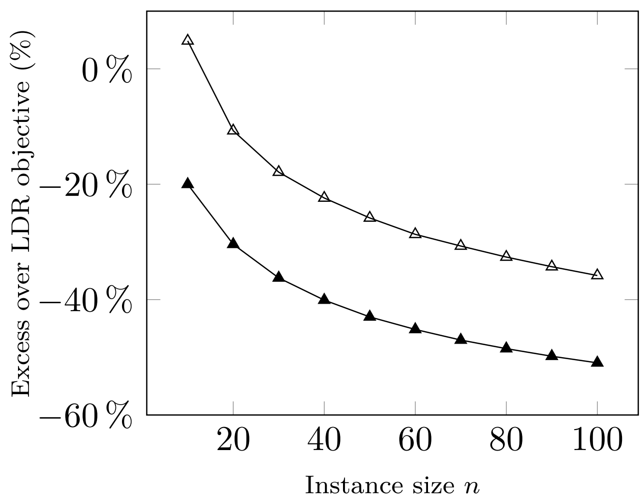


Figure 4.1 – Excess for two different decision rules.
−△− Piecewise linear decision rule (Ben-Tal et al. 2016)
−▲− Nonlinear decision rule (4.8)

Table 4.1 – Comparison of the average performance of the pure linear decision rules (LDR) from (4.2), the piecewise linear decision rule (PLDR) from Ben-Tal et al. (2016) and the nonlinear decision rules (NDR) from (4.8). (Excess in parenthesis)

n	LDR	PLDR	NDR
10	1.648	1.726 (**4.84**%)	1.318 (**-20.02**%)
20	2.359	2.105 (**-10.73**%)	1.641 (**-30.41**%)
30	2.909	2.388 (**-17.91**%)	1.855 (**-36.24**%)
40	3.388	2.629 (**-22.39**%)	2.030 (**-40.09**%)
50	3.811	2.826 (**-25.85**%)	2.172 (**-43.00**%)
60	4.194	2.991 (**-28.69**%)	2.299 (**-45.17**%)
70	4.547	3.150 (**-30.71**%)	2.409 (**-47.01**%)
80	4.876	3.286 (**-32.62**%)	2.511 (**-48.50**%)
90	5.186	3.407 (**-34.30**%)	2.603 (**-49.82**%)
100	5.48	3.517 (**-35.82**%)	2.687 (**-50.96**%)

We see that the nonlinear decision rules strictly outperforms the (already good performing) piecewise linear decision rules on the excess, as well as the pure linear decision rules on all instances. Note that our method with nonlinear decision rules is always guaranteed to give an objective value that is at least as good as solutions obtained with pure linear decision rules. The piecewise linear decision rules mostly outperform the linear decision rules, but this is not guaranteed as can be seen for instances of size $n = 10$, where the excess is positive. However, the model with piecewise linear decision rules is much faster to solve than any of the other methods. The small linear program resulting from the dominating set method was always solved within 0.3 seconds, even for $n = 100$, whereas the nonlinear decision rule (4.8) needed on average 58.4 seconds for $n = 100$ with a maximum solve time of 72.67 seconds. This is not much more than the solve times for the pure linear decision rules where on average 47.34 seconds were needed with a maximum of 53.13 seconds. Hence, the lower objective value obtained by using nonlinear decision rules only comes at small computational cost compared to the pure linear decision rules. Furthermore, if one is satisfied with a slightly worse objective value, then one could also stop the solver before reaching optimality. In that way, it is still possible to obtain solutions with nonlinear decision rules that are better than the linear decision rules or the piecewise linear decision rules.

4.7 Example 2: retailer-supplier flexible commitment contracts

4.7.1 The model

In the retailer-supplier flexible commitment contracts model the retailer makes commitments on order sizes for each period at the start of the planning horizon. At the beginning of each period the retailer is allowed to deviate from previous commitments in return for a certain penalty cost. In this way the retailer and supplier share the burden of uncertainty in the demand. The model we employ is taken as is from Ben-Tal et al. (2005). For completeness, we briefly describe the model here.

At the start of the planning horizon the retailer has an initial inventory of x_1. The retailer makes commitments $w_1, w_2, \dots, w_T$ representing the amount he is committed to order in each of the periods $1, \dots, T$. At the beginning of each period t, the retailer orders a quantity of $q_t(d)$ from the supplier at a unit cost of c_t. These decisions are adjustable and are allowed to depend on demand realizations from the past, i.e., on realizations of $d_1, \dots, d_{t-1}$. During each period the retailer has to order between a minimum of L_t and a maximum of U_t units. There are also lower bounds $\hat{L}_t$ and upper bounds $\hat{U}_t$ on the cumulative order quantities $\sum_{\tau=1}^{t} q_\tau$ in each period $t = 1, \dots, T$. The costs he incurs are holding (h_t) and backlog cost (p_t), penalty cost on deviating from the committed orders (α_t^+ and α_t^-) and penalties on deviations between successive commitments (β_t^+ and β_t^-). Any inventory that is remaining at the end of the horizon can be salvaged for s_{T+1} per unit. The final model we then obtain (described without uncertainty) is given below, see also equation (6) in Ben-Tal et al. (2005):

$$
\begin{aligned}
\min_{w_t, z_t, q_t, y_t, u_t, E} \quad & E \\
\text{s.t.} \quad & \sum_{t=1}^{T} (c_t q_t + y_t + u_t + z_t) \leq E \\
& L_t \leq q_t \leq U_t \qquad t = 1, \dots, T \\
& \hat{L}_t \leq \sum_{\tau=1}^{t} q_\tau \leq \hat{U}_t \qquad t = 1, \dots, T \\
& y_t \geq \bar{h}_t \left(x_0 + \sum_{\tau=1}^{t} q_\tau - \sum_{\tau=1}^{t} d_\tau \right) \qquad t = 1, \dots, T \\
& y_t \geq -p_t \left(x_0 + \sum_{\tau=1}^{t} q_\tau - \sum_{\tau=1}^{t} d_\tau \right) \qquad t = 1, \dots, T \\
& u_t \geq \alpha_t^+ (q_t - w_t), \; u_t \geq -\alpha_t^- (q_t - w_t) \qquad t = 1, \dots, T \\
& z_t \geq \beta_t^+ (w_t - w_{t-1}), z_t \geq -\beta_t^- (w_t - w_{t-1}) \qquad t = 2, \dots, T,
\end{aligned}
$$

where $\bar{h}_t = h_t$ for $t = 1, \ldots, T-1$ and $\bar{h}_T = h_T - s_{T+1}$ and w_0 are given. To model the holding, backlog and the penalty costs we introduced auxiliary variables $y_t(d)$, $u_t(d)$ and $z_t(d)$ for $t = 1, \ldots, T$. Since these are auxiliary variables (and not values that have to be set in each period like $q_t(d)$), we can allow these variables to depend on demand from both past and future (so $d_1, \ldots, d_T$) instead of only past demand realizations.

In the original paper by Ben-Tal et al. (2005) the main focus is on box uncertainty sets to model the uncertain behaviour of the demand. For box uncertainty sets the pure linear decision rules have spectacular performance. The authors showed numerically that linear decision rules are optimal among all decision rules. This model with box uncertainty has been studied by Bertsimas et al. (2010) and Iancu et al. (2013) whom show that the structure of the model almost theoretically guarantees optimality of linear decision rules for box uncertainty sets. The case is completely different for ellipsoidal uncertainty sets. Here even simpler inventory models considered by Gorissen and den Hertog (2013) already show that linear decision rules can be suboptimal.

For the uncertain demand we take $d_t = (1 + \frac{\rho}{100}\zeta)\bar{d}_t$, for $t = 1, \ldots, T$ with $\rho \in [0, 100]$ depicting the level of uncertainty and $\bar{d}_t$ the nominal demand. For the uncertainty set in which ζ resides we use the ellipsoidal uncertainty set (4.5) with $\Gamma = 1$. The data is taken as is from Ben-Tal et al. (2005, Table 1), which is for completeness given in Table 4.2.

4.7.2 Numerical results

We have compared the objective value using the decision rules from Chen and Zhang (2009) given by (4.21), the nonlinear decision rules with squares only (4.8) and the nonlinear decision rules that incorporates both absolute values and squares (4.19) from Section 4.5. We cannot use the piecewise linear decision rule from Ben-Tal et al. (2016), because of the restrictions that the method places on the structure of the problem. We use again the definition of excess as given in (4.23). The results for one dataset, $A12$ with uncertainty levels $\rho = 10\%, 20\%, \ldots, 100\%$ with radius $\Gamma = 1$ for the ellipsoidal uncertainty set, are given in Table 4.3. We have calculated the excess for each of the nonlinear decision rules for each run individually for all instances. We graphically depicted the average excess for dataset A12 in Figure 4.2 and also gave the average excess values in Table 4.3. The same figures and tables for the other datasets A10, D2 and W12 are given in Appendix 4.B.

Table 4.2 – Data sets from Ben-Tal et al. (2005).

parameter	A12	D2	W12	A10
T	12	12	12	10
x_1	57	100	0	18.895
w_0	12	100	100	19.9613
c_t	1.01	40	10	0.1818
h_t	0.3	2	2	4.6194
p_t	1.0	5	10	1.193
s_{T+1}	1.13	7	0	0.2195
α_t^+	0.43	2	10	1.204
α_t^-	0.58	3	10	2.11
β_t^+	0.37	1	10	0.4308
β_t^-	0.04	2	10	0.6278
L_t	44	50	0	14.9948
U_t	$76, 54, 66, 88, 68, 60$ $82, 53, 53, 78, 72, 63$	50	0	14.9948
$\hat{L}_t$	0	0	0	0
$\hat{U}_t$	814	∞	$200t$	∞
$\bar{d}_t$	64	100	100	46

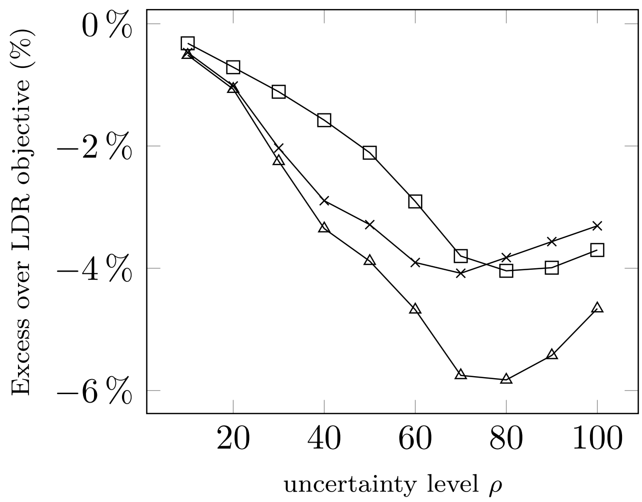


Figure 4.2 – Excess over objective value with pure linear decision rules on dataset A12 for different decision rules.
−□− Lifted decision rule (4.21) from Chen and Zhang (2009)
− × − Decision rules with squares (4.8)
−△− Decision rule with absolute values and squares (4.19).

Table 4.3 – Performance comparison of linear decision rules (LDR), the lifted decision rules (4.21) of Chen and Zhang (2009) (LDR-CH), the nonlinear decision rules with squares only (4.8) (NDR-SQ) and the further lifted decision rules with absolute values and squares (4.19) (NDR-ABSSQ). Results depicted are for dataset A12.

	Optimal objective value (excess in parentheses)			
ρ	LDR	LDR-CH	NDR-SQ	NDR-ABSSQ
10	821	819 (**-0.32**%)	817 (**-0.47**%)	817 (**-0.51**%)
20	864	858 (**-0.71**%)	855 (**-1.02**%)	855 (**-1.07**%)
30	918	907 (**-1.11**%)	899 (**-2.03**%)	897 (**-2.25**%)
40	981	966 (**-1.58**%)	953 (**-2.89**%)	948 (**-3.35**%)
50	1051	1029 (**-2.11**%)	1016 (**-3.29**%)	1010 (**-3.88**%)
60	1127	1094 (**-2.91**%)	1083 (**-3.91**%)	1074 (**-4.68**%)
70	1232	1186 (**-3.80**%)	1182 (**-4.08**%)	1161 (**-5.75**%)
80	1355	1301 (**-4.04**%)	1304 (**-3.82**%)	1276 (**-5.83**%)
90	1487	1428 (**-3.99**%)	1434 (**-3.56**%)	1407 (**-5.43**%)
100	1625	1565 (**-3.70**%)	1571 (**-3.31**%)	1549 (**-4.66**%)

We see improvement of several percentage points of using nonlinear decision rules over the pure linear decision rules. The nonlinear decision rule (4.8) with squares mostly outperforms the decision rule from Chen and Zhang (2009). The nonlinear decision rule (4.19), that has both the absolute values and squares of the uncertain parameter, strictly outperforms both. Note that the improvement also very much depends on the data, where the most improvement is made for dataset $A10$ and virtually none for dataset $D2$ as can be seen in Appendix 4.B. For all datasets, except $D2$, the nonlinear decision rules with both absolute values and squares gives a few percentages more improvement over the pure linear decision rules than the lifting from Chen and Zhang (2009).

4.8 Conclusion

In this chapter we introduced nonlinear decision rules for some convex uncertainty sets such as ellipsoidal and p-norm uncertainty sets. We show that the models with nonlinear decision rules are equivalent to linear decision rules on a lifted uncertainty set. Both theoretically and numerically we show that our nonlinear decision rules outperform the pure linear decision rules, as well as other decision rules from the literature.

For further research on a theoretical level it would be interesting to show bounds as in Theorem 4.1 for ϕ-divergence sets. For implementation considerations it would be convenient for end-users to have automatized software packages that can construct the tractable robust counterpart (4.13). Note that there are already several packages that can do this for adjustable robust optimization with pure linear decision rules such as ROME (Goh and Sim 2011). Therefore, the extension to our nonlinear decision rules should be straightforward. Furthermore, the lifting procedure of uncertainty sets could be automatized by detecting the convex structure and introducing the auxiliary variables. This can be achieved using disciplined convex programming, a method that is used to verify convexity of functions by breaking functions down to atoms, see Grant et al. (2006). In our case, the atoms would be the seperate functions $g_i(\zeta_i)$ in (4.4).

4.A Derivation of robust counterparts

Proposition 4.3 $(x, \bar{y}, Y, W)$ *is feasible (and optimal) for* (4.9) *with right-hand side uncertainty and the lifted ellipsoidal ($p = 2$) uncertainty set* (4.12) *if and only if there exist* $\lambda \in \mathbb{R}^{n\times m}$, $\psi \in \mathbb{R}^m$ *and* $t \in \mathbb{R}^{n\times m}$ *such that* $(x, \bar{y}, Y, W, \psi, \lambda, t)$ *is feasible (and optimal) for*

$$\begin{aligned}
\min_{x,\bar{y},Y,W,\psi,\lambda,t} \quad & c^\top x \\
s.t. \quad & A_j x + B_j \bar{y} - \sum_{i=1}^{n} t_{i,j} - \Gamma\psi_j \geq d_j \qquad j = 1,\dots,m \\
& \left\| \left(2\left(Y_i^\top B_j - D_{j,i}\right), 4\lambda_{i,j} - t_{i,j}\right) \right\|_2 \leq 4\lambda_{i,j} + t_{i,j} \qquad \forall i = 1,\dots,n, \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = 1,\dots,m \\
& -W_i^\top B_j + \lambda_{i,j} \leq \psi_j \qquad \forall i = 1,\dots,n,\ j = 1,\dots,m \\
& \lambda_{i,j} \geq 0 \qquad \forall i = 1,\dots,n,\ j = 1,\dots,m \\
& \psi_j \geq 0 \qquad \forall j = 1,\dots,m \\
& x \in \mathcal{X}.
\end{aligned} \tag{4.24}$$

Proof. Since model (4.9) is a linear robust optimization model we can replace $\mathcal{V}^{ext}$ by its convex hull $\mathcal{V}$ given in (4.11) derived in Proposition 4.1. To use the methods in Ben-Tal et al. (2015), we have to derive the dual of the support function of $\mathcal{V}$, which is defined as $\delta^*\left((s,z) \mid \mathcal{V}\right) := \max_{(\zeta,u)\in\mathcal{V}} \left\{s^\top\zeta + z^\top u\right\}$. By duality we derive for the support function:

$$\begin{aligned}
\delta^*\left((s,z)\mid\mathcal{V}\right) &= \max_{\zeta,u}\left\{s^\top\zeta+z^\top u \;\middle|\; \zeta_i^2\le u_i,\ i=1,\dots,n,\ \sum_{i=1}^n u_i\le\Gamma\right\}\\
&= \min_{\lambda,\psi}\left\{\sum_{i=1}^n \frac{s_i^2}{4\lambda_i}+\psi\Gamma \;\middle|\; \psi\ge 0,\ \lambda_i\ge 0,\ z_i+\lambda_i-\psi\le 0,\ i=1,\dots,n\right\}\\
&= \min_{\lambda,\psi,t}\Bigg\{\sum_{i=1}^n t_i+\psi\Gamma \;\Bigg|\; \psi\ge 0,\ \|(2s_i,4\lambda_i-t_i)\|\le 4\lambda_i+t_i,\\
&\qquad\qquad \lambda_i\ge 0,\ z_i+\lambda_i-\psi\le 0,\ i=1,\dots,n\Bigg\},
\end{aligned}$$

where for the second equality we used duality for quadratic programming. For the last equality we used the equivalent second-order cone formulation from Lobo et al. (1998):

$$\alpha^2\le\beta\gamma \;\Leftrightarrow\; \|(2\alpha,\beta-\gamma)\|_2\le\beta+\gamma.$$

By Ben-Tal et al. (2015, Theorem 1) we have that the j-th constraint, $j=1,\dots,m$ of (4.9)

$$A_jx+B_j\left(\bar{y}+Y\zeta+Wu\right)\ge D_j\zeta+d_j,$$

is satisfied if and only if the following constraint is satisfied

$$A_jx+B_j\bar{y}-\delta^*\left(\left(-Y^\top B_j+D_j,-W^\top B_j\right)\;\middle|\;\mathcal{V}\right)\ge d_j,$$

or equivalently, the j-th constraint is satisfied for $(x,\bar{y},Y,W)$ if and only if there exist $\lambda_{i,j},t_{i,j}$, $i=1,\dots,n$ and ψ_j such that

$$\begin{aligned}
&A_jx+B_j\bar{y}-\sum_{i=1}^n t_{i,j}-\Gamma\psi_j\ge d_j\\
&\left\|\left(2\left(Y_i^\top B_j-D_{j,i}\right),4\lambda_{i,j}-t_{i,j}\right)\right\|_2\le 4\lambda_{i,j}+t_{i,j} \qquad \forall i=1,\dots,n\\
&-W_i^\top B_j+\lambda_{i,j}\le\psi_j \qquad \forall i=1,\dots,n\\
&\lambda_{i,j}\ge 0 \qquad \forall i=1,\dots,n\\
&\psi_j\ge 0.
\end{aligned}$$

If we replace all $j=1,\dots,m$ constraints in this way, we end up with (4.24). ■

We now consider the more general case, where $A(\zeta)$ is allowed to depend affine on ζ:

$$A(\zeta)=A^0+\sum_{i=1}^n A^i\zeta_i.$$

Here the matrices $A^0, A^1, \ldots, A^n$ are elements of $\mathbb{R}^{m \times n_x}$. We denote the j-th row of matrix A^i by $(A^i)_j$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The robust counterpart is presented below for general convex functions g_i in the uncertainty set (4.7). The robust counterpart uses the conjugate functions of g_i, which are defined as:

$$(g_i)^*(\mu) = \sup_{\zeta_i \in \mathbb{R}} \left\{ \mu^\top \zeta_i - g_i(\zeta_i) \right\},$$

for all $i = 1, \ldots, n$. Note that the conjugate function is a convex function.

Proposition 4.4 $(x, \bar{y}, Y, W)$ *is feasible (and optimal) for* (4.9) *with the lifted uncertainty set* (4.11) *if and only if there exist* $\lambda, S \in \mathbb{R}^{n \times m}$ *and* $\psi \in \mathbb{R}^m$ *such that* $(x, \bar{y}, Y, W, \lambda, S, \psi)$ *is feasible (and optimal) for*

$$\begin{aligned}
\min_{x,\bar{y},Y,W,\lambda,S,\psi} \quad & c^\top x \\
s.t. \quad & (A^0)_j x + B_j \bar{y} - \sum_{i=1}^n \lambda_{i,j} (g_i)^* \left(\frac{S_{i,j}}{\lambda_{i,j}} \right) - \psi_j \Gamma \geq d_j \quad \forall j = 1, \ldots, m \\
& S_{i,j} = -(A^i)_j x - Y_i^\top B_j + D_{j,i} \qquad \forall i = 1, \ldots, n, \ j = 1, \ldots, m \\
& -W_i^\top B_j + \lambda_{i,j} \leq \psi_j \qquad \forall i = 1, \ldots, n, \ j = 1, \ldots, m \\
& \lambda_{i,j} \geq 0 \qquad \forall i = 1, \ldots, n, \ \forall j = 1, \ldots, m \\
& \psi_j \geq 0 \qquad \forall j = 1, \ldots, m. \\
& x \in \mathcal{X}.
\end{aligned} \tag{4.25}$$

Proof. Since model (4.9) is a linear robust optimization model we can replace $\mathcal{V}^{ext}$ by its convex hull $\mathcal{V}$ given in (4.11) derived in Proposition 4.1. To use the methods in Ben-Tal et al. (2015), we have to derive the dual of the support function of $\mathcal{V}$, which is defined as $\delta^* ((s, z) \mid \mathcal{V}) := \max_{(\zeta,u) \in \mathcal{V}} \left\{ s^\top \zeta + z^\top u \right\}$. By duality we derive for the support function:

$$\begin{aligned}
\delta^* ((s, z) \mid \mathcal{V}) &= \max_{\zeta,u} \left\{ s^\top \zeta + z^\top u \,\middle|\, g_i(\zeta_i) \leq u_i, \ i = 1, \ldots, n, \ \sum_{i=1}^n u_i \leq \Gamma \right\} \\
&= \min_{\lambda,\psi} \left\{ \sum_{i=1}^n \lambda_i (g_i)^* \left(\frac{s_i}{\lambda_i} \right) + \psi \Gamma \,\middle|\, \psi, \lambda_i \geq 0, \ z_i + \lambda_i - \psi = 0, \ i = 1, \ldots, n \right\},
\end{aligned}$$

where for the second equality we used Fenchel duality, see Rockafellar (1970). By Ben-Tal et al. (2015, Theorem 1) we have that the j-th constraint, $j = 1, \ldots, m$ of (4.9)

$$\left(A^0 + \sum_{i=1}^n A^i \zeta_i \right) x + B_j \left(\bar{y} + Y \zeta + W u \right) \geq D_j \zeta + d_j,$$

is satisfied if and only if there exist $s_{i,j}$, $i = 1, \dots, n$ such that the following constraint is satisfied:

$$\begin{aligned}
&(A^0)_j x + B_j \bar{y} - \delta^* \left(\left(S_{i,j}, , -W^\top B_j \mid \mathcal{V} \right) \;\Big|\; \mathcal{V} \right) \geq d_j \\
&S_{i,j} = -(A^i)_j x - Y_i^\top B_j + D_j \qquad i = 1, \dots, n,
\end{aligned}$$

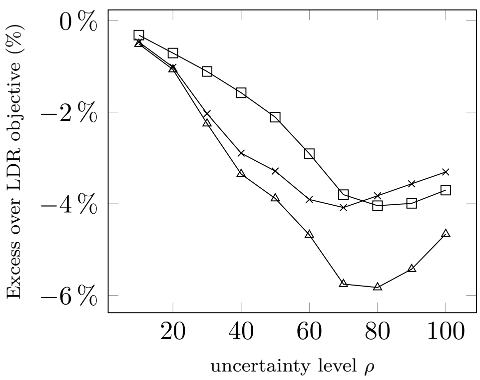
or equivalently, the j-th constraint is satisfied if and only if there exist $\lambda_{i,j}$, $s_{i,j}$, $i = 1, \dots, n$ and ψ_j such that

$$\begin{aligned}
&(A^0)_j x + B_j \bar{y} - \sum_{i=1}^{n} \lambda_{i,j} (g_i)^* \left(\frac{S_{i,j}}{\lambda_{i,j}} \right) - \psi_j \Gamma \geq d_j \\
&S_{i,j} = -(A^i)_j x - Y_i^\top B_j + D_j \qquad \forall i = 1, \dots, n \\
&- W_i^\top B_j + \lambda_{i,j} \leq \psi_j \qquad \forall i = 1, \dots, n \\
&\lambda_{i,j} \geq 0 \qquad \forall i = 1, \dots, n \\
&\psi_j \geq 0.
\end{aligned}$$

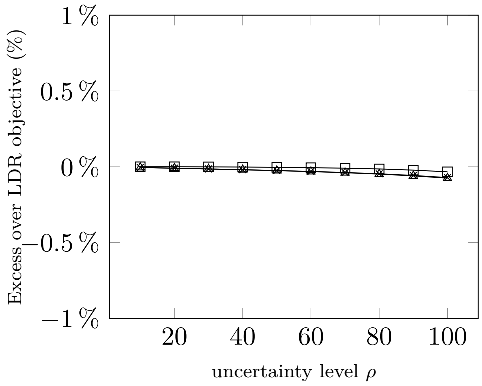
If we replace all $j = 1, \dots, m$ constraints in this way, we end up with (4.25). ■

Note that model (4.25) includes the terms $\lambda_{i,j}(g_i)^* \left(\frac{(A^i)_j x + Y_i^\top B_j - D_{j,i}}{\lambda_{i,j}} \right)$, with $\lambda_{i,j} \geq 0$, $i = 1, \dots, n$ and $j = 1, \dots, m$ which involve the perspective functions of the conjugate functions $(g_i)^*$ for all $i = 1, \dots, n$. The perspective function of a convex function is again convex, see Dacorogna and Maréchal (2007), making model (4.25) also convex (note the greater-than sign in the first m constraints).

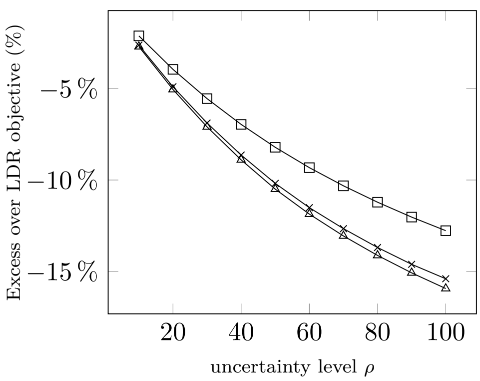
4.B More results on the retailer-supplier flexible commitment example



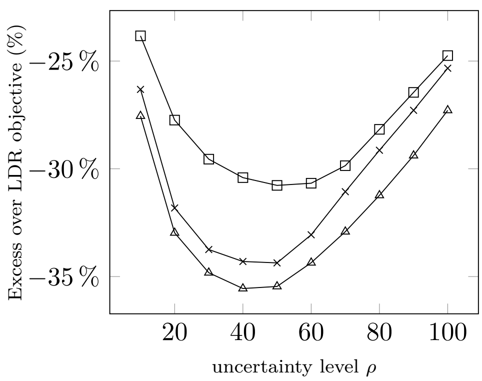
(a) Dataset A12.



(b) Dataset D2.



(c) Dataset W12.



(d) Dataset A10.

Figure 4.3 – Excess over objective value with pure linear decision rules for the four datasets for different decision rules.
−□− Lifted decision rule (4.21) from Chen and Zhang (2009)
− × − Decision rules with squares (4.8)
−△− Decision rule with absolute values and squares (4.19).

Table 4.4 – Performance comparison of linear decision rules (LDR), the lifted decision rules (4.21) of Chen and Zhang (2009) (LDR-CH), the nonlinear decision rules with squares only (4.8) (NDR-SQ) and the further lifted decision rules with absolute values and squares (4.19) (NDR-ABSSQ). Results depicted are for dataset A12, D2, W12 and A10.

(a) Dataset A12

	Optimal objective value (excess in parentheses)			
ρ	LDR	LDR-CH	NDR-SQ	NDR-ABSSQ
10	821	819 (**-0.32**%)	817 (**-0.47**%)	817 (**-0.51**%)
20	864	858 (**-0.71**%)	855 (**-1.02**%)	855 (**-1.07**%)
30	918	907 (**-1.11**%)	899 (**-2.03**%)	897 (**-2.25**%)
40	981	966 (**-1.58**%)	953 (**-2.89**%)	948 (**-3.35**%)
50	1051	1029 (**-2.11**%)	1016 (**-3.29**%)	1010 (**-3.88**%)
60	1127	1094 (**-2.91**%)	1083 (**-3.91**%)	1074 (**-4.68**%)
70	1232	1186 (**-3.80**%)	1182 (**-4.08**%)	1161 (**-5.75**%)
80	1355	1301 (**-4.04**%)	1304 (**-3.82**%)	1276 (**-5.83**%)
90	1487	1428 (**-3.99**%)	1434 (**-3.56**%)	1407 (**-5.43**%)
100	1625	1565 (**-3.70**%)	1571 (**-3.31**%)	1549 (**-4.66**%)

(b) Dataset D2

	Optimal objective value (excess in parentheses)			
ρ	LDR	LDR-CH	NDR-SQ	NDR-ABSSQ
10	38427	38427 (**-0.00**%)	38425 (**-0.00**%)	38425 (**-0.00**%)
20	39503	39503 (**-0.00**%)	39499 (**-0.01**%)	39499 (**-0.01**%)
30	40580	40580 (**-0.00**%)	40574 (**-0.01**%)	40574 (**-0.01**%)
40	41657	41656 (**-0.00**%)	41649 (**-0.02**%)	41649 (**-0.02**%)
50	42734	42733 (**-0.00**%)	42724 (**-0.02**%)	42723 (**-0.03**%)
60	43812	43809 (**-0.01**%)	43799 (**-0.03**%)	43798 (**-0.03**%)
70	44890	44886 (**-0.01**%)	44873 (**-0.04**%)	44873 (**-0.04**%)
80	45970	45963 (**-0.01**%)	45949 (**-0.05**%)	45948 (**-0.05**%)
90	47051	47040 (**-0.02**%)	47024 (**-0.06**%)	47022 (**-0.06**%)
100	48133	48117 (**-0.03**%)	48099 (**-0.07**%)	48097 (**-0.08**%)

(c) Dataset W12

	Optimal objective value (excess in parentheses)			
ρ	LDR	LDR-CH	NDR-SQ	NDR-ABSSQ
10	12935	12662 (**-2.12**%)	12596 (**-2.63**%)	12585 (**-2.71**%)
20	13871	13323 (**-3.95**%)	13191 (**-4.90**%)	13171 (**-5.05**%)
30	14806	13985 (**-5.55**%)	13787 (**-6.88**%)	13756 (**-7.09**%)
40	15742	14646 (**-6.96**%)	14382 (**-8.63**%)	14341 (**-8.90**%)
50	16677	15308 (**-8.21**%)	14979 (**-10.18**%)	14926 (**-10.50**%)
60	17612	15970 (**-9.32**%)	15585 (**-11.51**%)	15523 (**-11.86**%)
70	18548	16634 (**-10.32**%)	16200 (**-12.66**%)	16126 (**-13.06**%)
80	19483	17299 (**-11.21**%)	16816 (**-13.69**%)	16732 (**-14.12**%)
90	20420	17964 (**-12.03**%)	17437 (**-14.61**%)	17341 (**-15.07**%)
100	21357	18630 (**-12.77**%)	18069 (**-15.40**%)	17955 (**-15.93**%)

(d) Dataset A10

	Optimal objective value (excess in parentheses)			
ρ	LDR	LDR-CH	NDR-SQ	NDR-ABSSQ
10	199	151 (**-23.83**%)	147 (**-26.32**%)	144 (**-27.55**%)
20	312	226 (**-27.74**%)	213 (**-31.82**%)	209 (**-32.97**%)
30	443	312 (**-29.55**%)	293 (**-33.75**%)	289 (**-34.82**%)
40	579	403 (**-30.41**%)	380 (**-34.30**%)	373 (**-35.56**%)
50	718	497 (**-30.77**%)	471 (**-34.37**%)	463 (**-35.46**%)
60	857	594 (**-30.67**%)	574 (**-33.06**%)	563 (**-34.36**%)
70	998	700 (**-29.85**%)	688 (**-31.06**%)	669 (**-32.92**%)
80	1138	818 (**-28.17**%)	807 (**-29.14**%)	783 (**-31.23**%)
90	1279	941 (**-26.45**%)	930 (**-27.28**%)	903 (**-29.39**%)
100	1421	1069 (**-24.75**%)	1061 (**-25.33**%)	1033 (**-27.29**%)

CHAPTER 5

The impact of the existence of multiple adjustable robust solutions

5.1 Introduction

In Ben-Tal et al. (2004) the Robust Optimization (RO) methodology is extended to multi-stage problems. The proposed Adjustable Robust Optimization (ARO) techniques appeared to be very effective to solve uncertain multi-stage optimization problems. This first paper on ARO has been cited more than 800 times already, and the ARO methodology has been applied to a wide variety of problems (see e.g. the survey papers Bertsimas et al. (2011b) and Gabrel et al. (2014b)). Recently, it was shown that (A)RO problems may have multiple optimal solutions, and that not all of these solutions are Pareto robustly optimal (Iancu and Trichakis 2013). A solution is called Pareto robustly optimal if there is no other robustly feasible solution that has better objective value for at least one scenario, and for all other scenarios in the uncertainty set the objective value is not worse.

In this note we show that the ARO model of the production-inventory problem in Ben-Tal et al. (2004), which is the seminal work on ARO, also has multiple optimal robust solutions. Although in robust optimization one operates in a distribution-free environment, an often used performance measure is the mean objective value, which is evaluated posteriorly assuming some information on the distribution of the parameters. For the cases considered in Ben-Tal et al. (2004), we show that among the optimal robust solutions, the difference in mean objective value can be as much as 21.9% and for individual realizations the difference can be up to 59.4%. This underlines the importance of the message in Iancu and Trichakis (2013) that ARO problems may have multiple optimal robust solutions. In such cases one can often find optimal robust solutions that are much better with respect to the mean objective value than solutions that were initially found. We also refer to Delage and Iancu (2015, p.11), where another discussion is presented that also particularly focussed on multi-stage problems.

We also extend the experiments performed in Ben-Tal et al. (2004) by including a folding horizon approach. In a folding horizon approach the model is re-optimized in each period using the available information at that point of time and only the decisions for the current time are implemented. Using this approach we find that there are still multiple optimal robust solutions, but the differences in mean costs diminish. This is mainly due to the fact that the here-and-now decisions are unique in almost all periods. As a last experiment, we also analyze the model and solutions we found when replacing the worst-case objective by an expected value objective. For the expected value objective we find that, for the seminal production-inventory problem considered here, the solution is unique.

In the second part of this note we discuss several important implications for practical ARO. The first implication is that, by ignoring the possibility of multiple solutions, one can incorrectly conclude that the ARO solution is not better than the RO solution, or even incorrectly conclude that ARO is (much) better than RO. The second implication is that even in cases where it is a priori known that RO and ARO are equivalent, i.e., they have the same worst-case optimal objective value, one cannot conclude that there is no value in using ARO. This is because in many cases there are ARO solutions that give much better solutions for the mean costs. The third implication is that even in cases where affine decision rules are (nearly) optimal, i.e., the optimal robust objective value cannot be improved by using nonlinear decision rule, one cannot conclude that there is no value in using nonlinear decision rules. Such a conclusion might be wrong, since nonlinear decision rules may yield much better solutions for the expected objective value. These implications are illustrated by using both the production-inventory example from Ben-Tal et al. (2004) and two toy examples.

Our aim is to convince users of ARO that one should always check for the existence of multiple solutions. In many papers on ARO it is not reported that one checked for possible existence of multiple solutions. These papers run the risk that much better solutions could have been found, or even that wrong conclusions have been drawn. For example, researchers who use the same production-inventory example as in the seminal work by Ben-Tal et al. (2004) to test new ARO methods, should be aware of the fact that this problem has many optimal robust solutions with big differences in mean costs.

5.2 Multiple adjustable robust solutions

To illustrate the implications of multiple adjustable robust solutions we use three problems. The first problem is the production-inventory problem by Ben-Tal et al. (2004) in its original setting. The second problem is an illustrative toy example where the existence of multiple solutions is more directly visible. The last toy problem we investigate is a two-stage facility location problem. For all models we study both the impact in a folding and in a non-folding horizon approach.

5.2.1 Production-inventory model by Ben-Tal et al. (2004)

We have repeated the experiments for the production-inventory problem by Ben-Tal et al. (2004). All solutions are obtained using the commercial solver Gurobi 6.0 (Gurobi Optimization 2015) programmed in the YALMIP language (Löfberg 2004) in MATLAB. All options of Gurobi were left at their default values.

We have found three distinct optimal robust solutions for the original model by Ben-Tal et al. (2004, p. 369-370). All of these solutions are optimal in a robust sense, i.e. they have the same worst-case costs, but costs differ for individual (non worst-case) realizations of the demand. The first solution was obtained by just solving the original model with Gurobi. The average costs of this solution turned out to be much higher than the solution reported by Ben-Tal et al. (2004). The second solution is the solution that performs best on the mean costs among all optimal robust solutions. It can be found via the following two-step approach similar to the methods used by Iancu and Trichakis (2013) to find so-called Pareto robustly optimal solutions:

1. Solve the original model from Ben-Tal et al. (2004), which gives a solution with minimal worst-case costs.

2. Change the objective into minimizing the costs for the nominal demand trajectory. Furthermore, add a constraint that ensures that the worst-case costs do not exceed the costs found in Step 1.

The solution obtained after step two is the 'Best' solution, the one that performs best on the expected objective value among all optimal robust solutions that use linear decision rules, assuming that nominal demand is equal to the expected demand. This solution is also Pareto robust, as follows from Iancu and Trichakis (2013, Corollary 1), because the model with linear decision rules only has constraints that are linear. The third solution is found by changing the objective in the second step into *maximizing* costs for the nominal demand trajectory. This we call the 'Worst' solution. Without the two-step approach, and some bad luck, one could have obtained this

Table 5.1 – Performance of the Best, First and Worst optimal robust solutions.

		Best sol.		First sol.				Worst sol.			
Uncertainty level	WC costs	Mean	Std	Mean	Std	Performance gap Mean	Performance gap Max	Mean	Std	Performance gap Mean	Performance gap Max
2.5%	35105	33932	178	35105	0	**3.5%**	7.2%	35105	0	**3.5%**	7.2%
5%	36389	34073	350	35953	142	**5.5%**	11.8%	36389	0	**6.8%**	14.6%
10%	38990	34416	691	38136	232	**10.8%**	24.9%	38990	0	**13.3%**	30.7%
20%	44273	35077	1373	40174	696	**14.5%**	39.4%	42766	315	**21.9%**	59.4%

solution as a 'First' solution, i.e. by solving the original problem formulation. The performances of these three optimal robust solutions are given in Table 5.1. The first column states the uncertainty level, for which we used the same levels as in Ben-Tal et al. (2004). If the level of uncertainty is 2.5%, then this indicates that in each period the realized demand could be up to 2.5% higher or lower than the nominal demand. The three solutions are all robustly optimal, so they have the same worst-case costs (WC costs). For each of those solutions we have determined the mean costs and the standard deviation. In Ben-Tal et al. (2004) the mean costs were approximated using 100 simulated demand trajectories drawn from a uniform distribution. The mean costs can also be determined exactly since the objective is linear in the uncertain demand. For the mean costs comparison we assume, as in the original paper, that the mean demand is given by the nominal demand scenario. The standard deviation was derived using the second moment of the uniform distribution, the distribution that was also used in the seminal paper by Ben-Tal et al. (2004) to sample the scenarios to calculate average costs.

As is clear from Table 5.1, the performances of the three solutions differ significantly. For both the 'First' solution and the 'Worst' solution we give the mean and maximum *performance gap*. The mean performance gap is just the percentage increase of the mean costs compared to the mean costs of the 'Best' solution. The maximum performance gap is the single demand trajectory that results in the largest difference in costs between the 'Best' solution and the 'Worst' (or 'First') solution. To explain how this gap is calculated, we determine the costs for the 'Worst' and the 'Best' solution, when trajectory $\boldsymbol{d}$ realizes, by respectively $OPT_W(\boldsymbol{d})$ and $OPT_B(\boldsymbol{d})$. These costs are linear in demand $\boldsymbol{d}$ because the original objective is linear, fixed recourse and we use linear decision rules. The maximum performance gap for the 'Worst' solution is given by

$$\max_{\boldsymbol{d}\in\mathcal{U}} \frac{OPT_W(\boldsymbol{d}) - OPT_B(\boldsymbol{d})}{OPT_B(\boldsymbol{d})},$$

where $\mathcal{U}$ is the box uncertainty set (defined by a set of linear constraints) used in this inventory problem. This is a linear-fractional maximization problem, which can be written as a linear optimization problem using the well-known Charnes-Cooper transformation (Charnes and Cooper 1962). The maximum performance gap for the 'First' solution is defined and determined analogously. The 'First' solution, which is the solution we obtained after solving the original LP problem with our solver, has mean costs of up to 14.5% higher than the mean costs for the best solution for a 20% uncertainty level. The 'Worst' solution has a performance gap of 21.9% for the same uncertainty level. If we compare the performance for individual realizations, we see that the costs can increase up to 39.4% and 59.4% for the 'First' and 'Worst' solutions, respectively. For uncertainty levels up to 10% the mean costs for the 'Worst' solution are equal to the worst-case costs, meaning that the worst-case costs are attained in every single scenario. Finally, as reported by Ben-Tal et al. (2004), only for an uncertainty level of 2.5% one can find a feasible nonadjustable solution implying that production levels in each period must be determined at the beginning of the planning horizon. The mean costs of 35279 for the nonadjustable solution are only slightly higher than the mean costs for the adjustable 'Worst' solution. Note that in the nonadjustable case there is no uncertainty in the objective, hence the mean costs are equal to the worst-case costs.

The mean costs of the solution reported by Ben-Tal et al. (2004), where no use of a two-step approach was reported, coincides with the performance of our 'Best solution'. We have tried various settings for our solver to see whether we could also replicate their good result as a 'First' solution. We tried both primal/dual simplex methods, interior point methods and a mixture of both in Gurobi. We have also solved the model for each of these options with crossover either enabled or disabled. If the crossover option is enabled, then the solver will push a solution in the optimal facet to a basic solution. None of these alterations led to a solution that was considerably better than our 'First' solution depicted in Table 5.1.

5.2.2 Folding horizon versus non-folding horizon

One might wonder whether the same differences in mean costs still exist if a so-called folding horizon (FH) is used. In a folding horizon approach the model is re-optimized at each period using the available information at that point of time and only the decisions for the current time are implemented. This is done for each period t starting from the first period until the end of the planning horizon. Using this folding horizon approach we again compared solutions that used the two-step approach in each step (Best FH solution), without a two-step approach (First FH solution) and when the two-step approach was used when maximizing for nominal demand in the second step (Worst FH solution). An exact calculation of the mean costs and the standard devia-

Table 5.2 – Performance of the best, first obtained and worst optimal robust solutions using the folding horizon approach.

	Best FH sol.		First FH sol.				Worst FH sol.			
Uncertainty level	Mean	Std	Mean	Std	Performance gap Mean	Performance gap Max	Mean	Std	Performance gap Mean	Performance gap Max
2.5%	33909	179	33912	179	**0.0%**	0.1%	33912	178	**0.0%**	0.1%
5%	34057	330	34061	328	**0.0%**	0.0%	34059	328	**0.0%**	0.0%
10%	34327	676	34350	667	**0.1%**	0.7%	34351	666	**0.1%**	0.7%
20%	34495	1361	34517	1348	**0.1%**	0.6%	34532	1339	**0.1%**	0.6%

tion is not possible for this experiment. Therefore, we draw 100 demand trajectories independently and uniformly distributed in each period. These trajectories are used to approximate the mean costs and the standard deviation when using the folding horizon approach. Simulations were also used in Ben-Tal et al. (2004) to approximate the mean costs and the standard deviation for the non-folding horizon approach. The results are depicted in Table 5.2. We stress that this folding horizon approach was not used in Ben-Tal et al. (2004). Clearly, using the two-step approach does not yield significantly better results for the folding horizon approach. Often the resulting costs are the same for both approaches, but for one of our simulated realizations the extra costs incurred when not using the two-step approach is 0.7%. Even stronger, for each simulated demand trajectory, the costs when using the folding horizon approach (Best FH solution) were at most the costs of the "First FH" solution. Finally, note that the mean costs for the folding horizon solutions are not much lower than the mean costs of the 'Best' solution given in Table 5.1, meaning that there is not much additional gain by re-optimizing in each step as is done in the folding horizon approach. It is at a first glance surprising that the effect of having multiple optimal solutions diminishes when using a folding horizon approach. We found that this is mainly because the first stage decisions are unique for almost all time periods and in all simulated scenarios. The question whether or not the first stage decisions are unique can be answered by fixing the worst case costs in the first step, as in the usual two step approach, and then minimize or maximize the order quantity in the current time period. In this way we get, for each time period t, a lower and upper bound on the feasible first stage decisions. In Figure 5.1 we depict the differences between the maximum and the minimum for the 20% uncertainty level for one out of the three factories. The behaviour of the solutions depicted was observed for all other cases as well: the vast majority of the first-stage decisions are unique. We only witnessed non-unique optimal here-and-now decisions in time periods 6 and 18, depending on the factory (1, 2 or 3) considered.

Finally, we also investigate what happens if we optimize the expected objective value

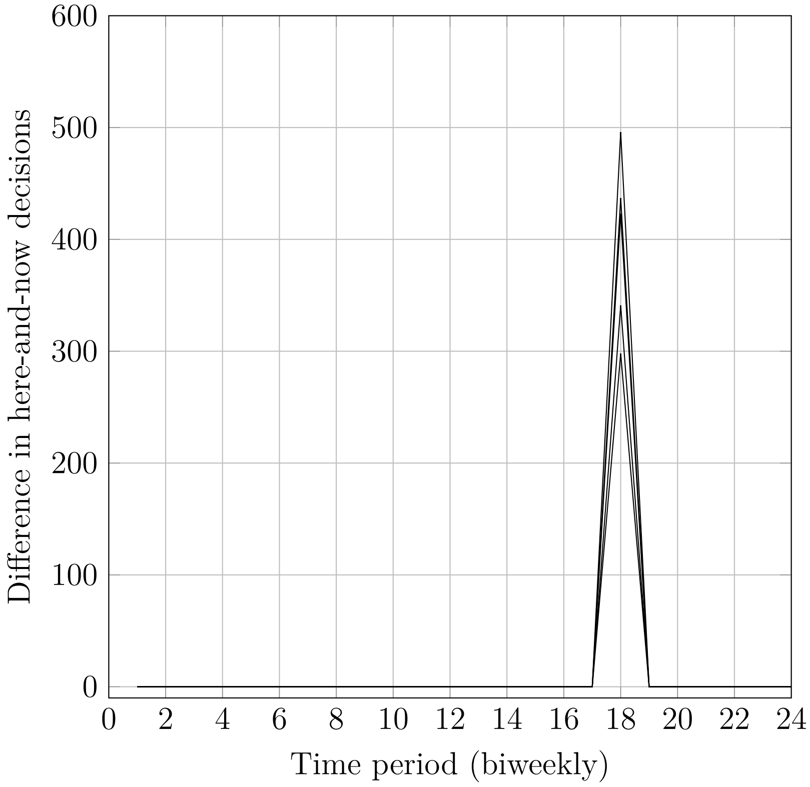


Figure 5.1 – Here-and-now decisions for factory 2 only differ in period 18 (5 scenarios depicted).

rather than the worst-case objective value in the non-folding horizon approach. This can be done at comparable computational costs, by replacing the maximization over all realizations in the objective by an objective that solely considers the nominal demand. This expected objective value was also used in Kuhn et al. (2011) to prove optimality of linear decision rules under stochastic and robust settings. The authors did not study the existence of multiple adjustable solutions. We stress that, although we now minimize an expected objective value, we still have a robust problem with 'hard' constraints, i.e., the constraints should be satisfied for any realization within the uncertainty set. The main difference with the two step approach is that we do not fix the worst-case objective value as we did in the second step. Arguably, this approach would make more sense in problems where the objective is a 'soft' criterion as opposed to the constraints which are typically 'hard' restrictions. When minimizing the expected objective value, the worst-case objective value is ignored. Hence, in principle, the worst-case costs could be very high. To find the worst-case objective value for a given linear decision rule, a posteriori, one can simply maximize the costs over all possible realizations within the uncertainty set. The results for the optimization problem, with the 'soft' expected objective value, but 'hard' constraints, are depicted in Table 5.3. First of all, we note that there is not much

Table 5.3 – Performance of the linear decision rule that minimizes the mean costs.

Uncertainty level	WC	Mean	Std
2.5%	35108	33919	178
5%	36412	34031	357
10%	39040	34311	708
20%	44298	35066	1375

difference between the mean costs and the worst-case costs with respect to the 'Best' robust solution given earlier in Table 5.1. There is only a very minor increase in worst-case costs and a very minor decrease in the mean costs. Hence, minimizing the mean costs yields a solution that has very similar costs to the costs of the solution obtained when minimizing worst-case costs. Second, there is no 'Best' and 'Worst' solution displayed in Table 5.3. This is because we found that the obtained solution is unique, so there does not exist a linear decision rule, with minimum mean costs, that has a different (neither better nor worse) guarantee on the worst-case objective value.

In the inventory model the decisions are made biweekly. Therefore, it makes sense to use a folding horizon approaches in this case. The impact of multiple adjustable robust solutions on the mean costs is negligible when we re-optimize. However, there might still be value in checking for multiple solutions in (non-)folding horizon approaches for inventory models and related multi-stage optimization models for the following reasons:

1. The non-folding horizon solution can be used as a backup solution in case of failure in hardware or software during the re-optimization steps. This is especially important in more critical multi-stage optimization systems such as power systems.

2. Re-optimization might take too much computation time or might not be possible at all. This happens in multi-stage optimization settings when periods follow up close in time, or when the solutions are implemented in low-end software systems. Examples of low-end computer systems are traffic light systems, that are not designed to solve the more computationally demanding optimization models.

Although for this inventory model the impact of the existence of multiple adjustable robust solutions on the mean costs seems to be negligible, there are other models where there could be a significant impact. This is illustrated by our toy examples in the next section.

5.2.3 Toy examples

Our first illustrative toy example is the following maximization problem:

$$\begin{aligned} \max_{x,y} \min_{a\in[0,1]} \quad & ax - y \\ \text{subject to} \quad & y + b^2 + b \geq 0 \qquad \forall b \in [0,1] \\ & 0 \leq x \leq 1. \end{aligned} \tag{toy-1}$$

Let us consider the case where both x and y are nonadjustable. We readily see that the worst-case objective value is 0 and the two solutions, $RO_1 = (1,0)$ and $RO_2 = (0,0)$, or any convex combination of these, are worst-case optimal. Without a two-step approach the solver is indifferent between all these optimal robust solutions since they all have optimal worst-case profits. The realized profits as a function of scenario (a,b) are respectively $p_{RO_1}(a,b) = a$ and $p_{RO_2}(a,b) = 0$ and the two-step approach yields solution RO_1.

Now suppose that y is adjustable and we restrict ourselves to linear decision rules (LDR). Then we find that linear decision rules $y(b) = -b$ or $y(b) = -\frac{1}{2}b$ are optimal in worst-case sense together with any nonadjustable x in $[0,1]$. For the first solution LDR_1 we take $(x,y) = (1,-b)$ and for the second solution $LDR_2 = (0, -\frac{1}{2}b)$. The profits of these solutions for scenario (a,b) are respectively $p_{LDR_1}(a,b) = a + b$ and $p_{LDR_2}(a,b) = \frac{1}{2}b$. Again, without a two-step approach the solver would be indifferent between these solutions since both have optimal worst-case objective value of 0. The two-step approach yields solution LDR_1.

Finally, we notice that the so-called *perfect hindsight* solution, where parameters a and b are known before deciding upon x and y, equals $(x,y) = (1, -b^2 - b)$ for any a, b in $[0,1]$. This perfect hindsight solution can also be obtained in the adjustable robust optimization model by allowing for nonlinear decision rules and setting $NDR_1 = (1, -b^2 - b)$. The profits for this nonlinear decision rule (NDR) are $p_{NDR_1}(a,b) = a + b^2 + b$ for scenario (a,b). Again, there are many more nonlinear decision rules that are optimal in worst-case sense, but have different mean profits. One example is $NDR_2 = (0, -\frac{1}{2}b^3)$ which yields profit $p_{NDR_2}(a,b) = \frac{1}{2}b^3$. All these results are summarized in Table 5.4.

In the table we use a uniform distribution to calculate the mean profits. For robust optimization one usually assumes to have only very crude information on the distribution function. Nevertheless, if we denote the mean profits of each solution by

Table 5.4 – Comparison of the different nonadjustable and adjustable solutions.

	RO_1	RO_2	LDR_1	LDR_2	NDR_1	NDR_2
Here-and-now x	1	0	1	0	1	0
Wait-and-see* y	0	0	$-b$	$-\frac{1}{2}b$	$-b^2-b$	$-\frac{1}{2}b^3$
Profits for scenario (a,b)	a	0	$a+b$	$\frac{1}{2}b$	$a+b^2+b$	$\frac{1}{2}b^3$
Worst-case profits	0	0	0	0	0	0
Mean profits (with unif. distr.)	$\frac{1}{2}$	0	1	$\frac{1}{4}$	$\frac{4}{3}$	$\frac{1}{8}$

*Note that for RO_1 and RO_2 the variable y is a here-and-now variable.

$\bar{p}_{RO_1}, \bar{p}_{RO_2}, \bar{p}_{LDR_1}, \bar{p}_{LDR_2}, \bar{p}_{NDR_1}$ and $\bar{p}_{NDR_2}$, then we have

$$\bar{p}_{NDR_1} > \bar{p}_{LDR_1} > \bar{p}_{RO_1} > \bar{p}_{LDR_2} > \bar{p}_{NDR_2} > \bar{p}_{RO_2}$$

for a large class of distribution functions. All these inequalities are valid if (1) not all probability mass of b lies on the extremes, i.e. $P(b = 0 \text{ or } b = 1) \neq 1$ and (2) the mean value of a and b is such that $\mathbb{E}(a) > \frac{1}{2}\mathbb{E}(b)$.

Note that for this toy example, contrary to the model from Ben-Tal et al. (2004), there could be a significant gain from the two-step method in the folding horizon approach. The variable x has to be chosen in the first step of the optimization. As we have seen, the optimal robust value is indifferent between any x in $[0,1]$. In the second step we shall always choose $y = -b^2 - b$. However, choosing $x = 0$ instead of $x = 1$ gives us a difference of a in the objective value. The two-step approach combined with the folding horizon approach returns the optimal (folding horizon) solution, which equals NDR_1.

Similar to our extended experiments for the numerical production-inventory example, we can also replace the worst-case objective by an expected value objective. Again, we find a unique solution when using an expected value objective to the following optimization model:

$$\begin{aligned} \max_{x,y} \quad & \mathbb{E}(a)x - \mathbb{E}\left(y(b)\right) \\ \text{subject to} \quad & y(b) + b^2 + b \geq 0 \qquad \forall b \in [0,1] \\ & 0 \leq x \leq 1. \end{aligned} \tag{toy-1-mean}$$

Now, if $\mathbb{E}(a) > 0$, then the solver returns the unique optimal $x = 1$. The only optimal (and unique) static and linear decision rules are given by $y(b) = 0$ and $y(b) = -b$, respectively. These are the same solutions as the best decision rules for the optimization problem with worst-case objective value. For the nonlinear decision rule we find that the optimal decision rule is

$$y(b) = -b - b^2 \qquad \text{(almost surely)}.$$

Our second toy example is a simple facility location problem with two facilities and a set of customers $\{1, \ldots, N\}$. The set of customers is such that the unit transportation costs from facility 1 and facility 2 to customer N are both equal to 10. All other customers are (much) closer to both facilities, but unit transportation costs from facility 2 are significantly smaller than from facility 1. This situation is depicted in Figure 5.2.

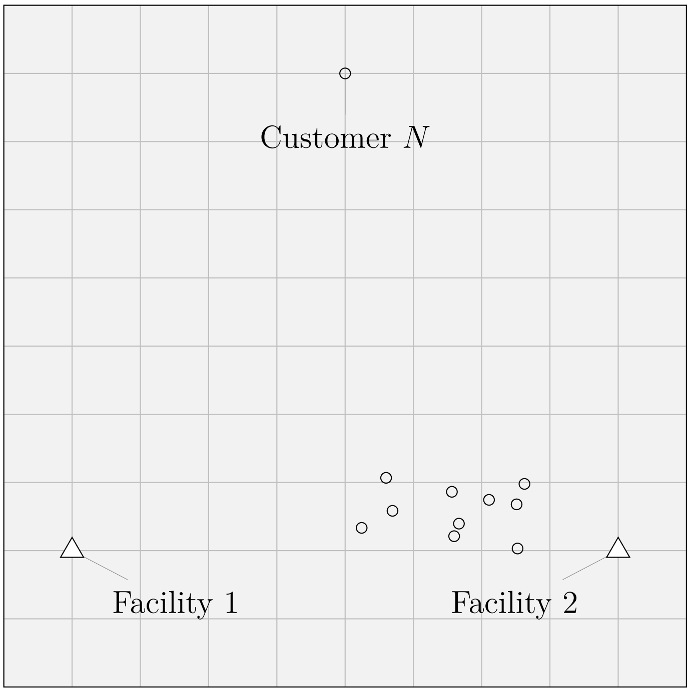


Figure 5.2 – Facility location problem with the most remote customer N at the same distance from both facilities. The two facilities are depicted by triangles, the customers by circles.

The demand of the customers is uncertain. In the entire network the demand is at most 1, but we do not know at which customers the demand will occur. We model this via the uncertainty set: $\mathcal{U} = \left\{d: \; d_i \geq 0 \; i = 1, \ldots, N, \; \sum_{i=1}^{N} d_i \leq 1\right\}$, where d_i denotes the uncertain demand of customer i. The facility location problem consists of two types of decisions, namely the decision to open facility 1 ($x_1 = 1$) or facility 2 ($x_2 = 1$) and the actual deliveries to the customers from the opened facility. Only one of the facilities may be opened. The total delivery to customer i from facility 1, respectively facility 2, is y_{1i} and y_{2i} and has unit costs c_{1i} and c_{2i}. The goal is to minimize the worst-case transportation costs, which is modeled as:

$$\begin{aligned}
\min_{x,y} \quad & \sum_{i=1}^{N} (c_{1i}y_{1i} + c_{2i}y_{2i}) \\
\text{subject to} \quad & y_{1i} + y_{2i} \geq d_i \qquad \forall i = 1, \ldots, N \quad \forall (d_1, d_2, \ldots, d_N) \in \mathcal{U} \\
& y_{1i} \leq x_1 \qquad \forall i = 1, \ldots, N \\
& y_{2i} \leq x_2 \qquad \forall i = 1, \ldots, N \\
& x_1 + x_2 \leq 1 \\
& x_1, x_2 \in \{0, 1\}.
\end{aligned} \tag{toy-2}$$

From Figure 5.2 it is clear that the transportation costs when facility 1 is opened are higher than when facility 2 is opened. The optimal perfect hindsight solution is to open facility 2 and transport exactly the requested demand $y_{2i}(d_i) = d_i$ to each customer. The costs for a particular demand realization $(d_1, d_2, \ldots, d_N)$ is then given by

$$\sum_{i=1}^{N} c_{2i}d_i.$$

The worst-case costs belonging to this solution are

$$\max_{(d_1,d_2,\ldots,d_N)\in\mathcal{U}} \sum_{i=1}^{N} c_{2i}d_i = c_{2N}.$$

In the nonadjustable robust model we decide upon all variables before we know the demand realization $d_1, \ldots, d_N$. The total demand in the network is 1, but all demand could occur at a single customer, so we have to transport one unit to each customer. Therefore, the first constraint in the robust model is equivalent to $y_{1i}+y_{2i} \geq 1$. Since $c_{1i} > c_{2i}$ for all customers $i = 1, \ldots, N-1$, the optimal solution is $x_1 = 0, x_2 = 1$ with $y_{1i} = 0, y_{2i} = 1$ for all $i = 1, \ldots, N$ and objective value $\sum_{i=1}^{N} c_{2i}$. The robust solution vastly overestimates the worst-case costs, but it does open facility 2. In the folding horizon approach, the transportation decisions are re-optimized and we obtain $y_{2i} = d_i$ with costs $\sum_{i=1}^{N} c_{2i}d_i$, which equal the costs in the perfect hindsight solution.

In the adjustable robust model there are multiple optimal solutions. In the first solution we open facility 1 and transport $y_{1i} = d_i$ to customer $i = 1, \ldots, N$. In the second solution we open facility 2 and transport $y_{2i} = d_i$ to each customer. Clearly, we obtain the same worst-case costs $c_{1N} = c_{2N}$ as in the perfect hindsight case. However, the costs when $(d_1, d_2, \ldots, d_N)$ realizes equals $\sum_{i=1}^{N} c_{1i}d_i$ and $\sum_{i=1}^{N} c_{2i}d_i$ respectively. If the expected demand is $d_i = \frac{1}{N}$ for all $i = 1, \ldots, N$, or any other scenario that does not place all probability mass on the demand realization with $d_N = 1$, then the two-step approach picks the solution that opens facility 2.

To conclude, in the first toy-example the here-and-now decisions matter in the folding horizon approach for the costs, but there is no impact of the existence of multiple here-and-now decisions on the choice of the optimal wait-and-see decision in the re-optimization step. In the second toy example we do see an impact: once the wrong facility is opened in the first stage, all demand has to be fulfilled from that location at high expected costs in the re-optimization step.

5.3 Implications for robust optimization

The inventory-production problem and the toy examples from the previous section allow us to present some important implications. First, if we analyze and compare the mean objective values of arbitrary optimal robust solutions for RO and ARO, then false conclusions can be drawn regarding the added value of ARO over RO. The mean objective value of an *arbitrary* optimal robust solution, obtained by solving the original RO or ARO problem formulations, might very well be much worse than the solution with best mean objective value among all optimal robust solutions. This best performing solution can be obtained by carrying out the two-step approach. In the production-inventory problem with uncertainty level 2.5%, the worst-case objective values of the RO and ARO solution are nearly the same: the difference is only 0.5%. If we compare the RO and ARO solutions on average costs, then the worst ARO solution is also only 0.5% better than the RO solution. The best ARO solution, however, is 3.5% better on average, which could be overlooked if the two-step approach is not carried out. For the 20% uncertainty level, the gap between the average performances of all optimal robust ARO solutions can be as much as 21.9%. The first toy example illustrates that an arbitrary ARO solution is not guaranteed to do better than a RO solution with respect to average performance. For instance, the average performance of robust solution RO_1 is better than the performance of ARO solution LDR_2. On the other hand, the optimal ARO solution LDR_1 is guaranteed to do better than any RO solution on the average performance. In our small facility location example we have seen that the robust solution results in a much higher objective value, but that it does open the best facility for folding horizon approaches. The linear decision rule on the other hand results in multiple optimal solutions which could lead to undesirable choices for opening the facilities. The two-step approach results in a solution that opens the cheapest facility, mimicking the solution of perfect hindsight.

Second, one might be inclined to jump to the conclusion that ARO can be safely ignored, when it is a priori known that ARO and RO are equivalent with respect to the *worst-case* objective value. One of the situations that we know where ARO is equivalent to RO is the case of constraint-wise uncertainty, see Ben-Tal et al. (2004, Theorem 2.1). However, the equivalence is not necessarily true for the mean objective value as well. Therefore, one should not ignore ARO for such problems. This is illustrated by the first toy example: the worst-case objective value is zero for both the RO and ARO solutions, but the mean objective values differ significantly.

Third, even if affine decision rules yield (near) optimal worst-case performance, nonlinear decision rules, such as quadratic decision rules, can yield much better performance on the mean objective value. Most applications of ARO restrict decision rules to affine functions, which is referred to as affinely adjustable robust optimization (Ben-Tal et al. 2004). Affine decision rules are known to perform optimal or nearly optimal in many situations (Ben-Tal et al. 2009; Bertsimas et al. 2010; Gounaris et al. 2013). However, once again, this observation is with respect to the worst-case objective value, and not for the mean objective value. This is illustrated by the first toy example. Here, the quadratic decision rule NDR_1 has the same worst-case objective value as any of the other decision rules, but the mean objective value is much better, and, in this particular case, even optimal for each scenario (Bellman optimal).

The encompassing recommendation that follows from these implications is that the two-step approach should always be conducted in any application of robust optimization. The two-step approach enables the optimizer to fully exploit the performance on the mean objective value of the solution, while guaranteeing no deterioration in the worst-case performance. This is especially relevant for ARO, where decision rules can be utilized to enhance the solution's performance in other than worst-case scenarios. We also recommend the use of the two step approach in folding horizon methods, but we do note that the impact of multiple solutions may be less severe.

CHAPTER 6

Robust optimization of uncertain multistage inventory systems with inexact data in decision rules

6.1 Introduction

With the uprise of Big Data, most of the currently available (theoretical or practical) methods for controlling a multi-stage production-inventory system, are using a "data-driven" approach. At each period t data in the future is treated as uncertain, while data from the past is considered known (certain). The Affinely Adjustable Robust Counterpart (AARC) method (Ben-Tal et al. 2004), which is the focus of this chapter, needs exact past demands to derive a decision, by inserting them in a linear decision rule. In reality, however, there is a strong evidence (see below) that even past data is far from being exact. For example, in inventory/production systems what is usually reported as a surrogate for the demand are sales, which then ignores lost sales due to excess demand.

In general, even when it seems that the full data on the uncertain demand is available at some stage, one cannot rely blindly on this information. Arguably, many developments in information technology have enabled firms to collect real-time data. However, despite these enormous developments in our Big Data era, poor data quality is still a big issue. In DeHoratius and Raman (2008) results of an empirical study are reported; they found that 65% of the inventory records were inaccurate, and "the value of the inventory reflected by these inaccurate records amounted to 28% of the total value of the expected on-hand inventory". In Redman (1998) it is estimated that 1–5% of data fields are erred, which led to a costs increase of 8–12% of revenue in some carefully studied cases, and to a consumption of 40–60% of the expenditure in service organizations. Haug et al. (2011) summarize the literature that deal with the big impact of poor data quality: "Less than 50% of companies claim to be very confident in the quality of their data", "75% of organizations have identified costs stemming from dirty data". See also Soffer (2010) for a general exploration of data inaccuracy in business processes. One paper that develops a method to handle inaccurate inventory records is by Kök and Shang (2007). Their approach assumes that the distribution of the errors (describing the inaccuracy) is known and that inspections can be made at certain costs to exactly observe these errors.

In this chapter we extend the AARC method to a method named Adjustable Robust Counterpart with decision rules based on Inexact Data (ARCID) that incorporate past data uncertainty while keeping the resulting (deterministic) robust counterpart tractable. This is our main contribution, and it is achieved using results and techniques from the current robust optimization arsenal.

We illustrate the benefits of the ARCID model by revisiting the inventory problem that was used in the first paper on ARO (Ben-Tal et al. 2004). Numerical results for this production-inventory problem show that if one neglects the inexact nature of the revealed data, then the resulting solution might violate the constraints in many scenarios. For our numerical example, violations occurred for up to 80% of the simulated demand trajectories. The ARCID model is able to avoid this severe infeasibility and produce more reliable solutions.

Although the focus of this chapter is on production-inventory problems, there are various other areas where our ARCID model could be used to solve uncertain multistage problems. For example, ARO techniques were used in facility location planning (Baron et al. 2011), flexible commitment models (Ben-Tal et al. 2005), portfolio optimization (Calafiore 2008; Calafiore 2009; Rocha and Kuhn 2012), capacity expansion planning (Ordóñez and Zhao 2007) and management of power systems (Guigues and Sagastizábal 2012; Ng and Sy 2014) among others. A more elaborate list of examples up to 2011 can be found in the aforementioned survey by Bertsimas et al. (2011b). We emphasize that our proposed ARCID framework remains applicable for multistage problems outside the realm of production-inventory planning.

The remainder of this chapter is organized as follows. In Section 6.2 we describe the adjustable robust models used in the literature. Section 6.3 then introduces the new ARCID models with inexact revealed data in the decision rules and derive tractable representations of the resulting optimization problems. Section 6.4 presents our production-inventory model and the corresponding ARCID model. The numerical results are given and analyzed in Section 6.5. Conclusions and some possible extensions are presented in Section 6.6. Throughout this chapter we use bold lowercase and upper-case letters for vectors and matrices, respectively, while scalars are printed in regular font.

6.2 Adjustable robust models

In the nonadjustable RC model all decisions are chosen prior to knowing the realization of the uncertain parameter. This can be very conservative in a dynamic setting where part of the variables can be chosen at a later stage when some information on the uncertain parameters is revealed. Suppose that $\mathbf{x} \in \mathbb{R}^n$ is a *here-and-now* decision and that we have an additional *wait-and-see* decision $\mathbf{y} \in \mathbb{R}^m$. This means that $\mathbf{x}$ has to be chosen prior to knowing any of the information on the uncertain parameters and $\mathbf{y}$ has to be chosen after some information is revealed. We start with the assumption that $\mathbf{y}$ is chosen after perfectly accurate information on $\boldsymbol{\zeta}$ has been revealed. The model with this underlying assumption is called the *adjustable robust optimization model* (ARO), where the variables $\mathbf{y}$ can adjust themselves to the revealed information. This model was introduced in Ben-Tal et al. (2004) similar to (1.6) introduced in Chapter 1:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{Z} \quad \exists \mathbf{y} \in \mathbb{R}^m : \quad (\mathbf{a}_i + \mathbf{A}_i \boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}_i^\top \mathbf{y} \le d_i \qquad \forall i = 1, \dots, J, \end{aligned} \tag{ARO}$$

where J is the number of constraints, $\mathbf{a}_i, \mathbf{c} \in R^n$, $\mathbf{A}_i \in \mathbb{R}^{n \times L}$, $\mathbf{b}_i \in \mathbb{R}^m$, and $d_i \in \mathbb{R}$. The uncertainty in our model is driven by the parameter $\boldsymbol{\zeta}$, which resides in a closed convex set $\mathcal{Z} \subset \mathbb{R}^L$. The parameter $\mathbf{a}_i$ is called the nominal value of the coefficients for $\mathbf{x}$ in the i-th constraint. This model can be readily extended to the case where d_i also depends on $\boldsymbol{\zeta}$. We can see $\mathbf{y}$ as a function, or decision rule, on the uncertain parameters since we have to assign a feasible value for each realization $\boldsymbol{\zeta}$. However, finding the optimal decision rule would involve optimizing over the class of all functions, which is in general intractable (in fact NP-hard as shown in Guslitzer (2002)). We restrict the functional dependence to linear decision rules for the wait-and-see decision:

$$\mathbf{y} = \mathbf{u} + \mathbf{V}\boldsymbol{\zeta},$$

where $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{V} \in \mathbb{R}^{m \times L}$ are new (here-and-now) decision variables that determine the affine dependence on the revealed value of the parameter $\boldsymbol{\zeta}$. Although the restriction from ‘any’ function to a linear decision rule might seem very severe, these linear decision rules appear to perform quite well in practice (Ben-Tal et al. 2004; Ben-Tal et al. 2005) and are even provably optimal in some cases (Bertsimas et al. 2011a; Bertsimas and Goyal 2012; Iancu et al. 2013; Gounaris et al. 2013). With this new so-called linear decision rule, the problem (ARO) can be written as

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u},\mathbf{V}} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{Z} : \quad (\mathbf{a}_i + \mathbf{A}_i \boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}_i^\top (\mathbf{u} + \mathbf{V}\boldsymbol{\zeta}) \le d_i \qquad \forall i = 1, \dots, J, \end{aligned} \tag{AARC}$$

Table 6.1 – Examples of uncertainty sets and their support functions.

Uncertainty set	$\mathcal{Z}$	$\delta^*(\boldsymbol{\nu}\vert\mathcal{Z})$
box	$\{\boldsymbol{\zeta} : \vert\vert\boldsymbol{\zeta}\vert\vert_\infty \leq \alpha\}$	$\alpha\vert\vert\boldsymbol{\nu}\vert\vert_1$
ball	$\{\boldsymbol{\zeta} : \vert\vert\boldsymbol{\zeta}\vert\vert_2 \leq \alpha\}$	$\alpha\vert\vert\boldsymbol{\nu}\vert\vert_2$
polyhedral	$\{\boldsymbol{\zeta} : \mathbf{b} - \mathbf{B}\boldsymbol{\zeta} \geq 0\}$	$\begin{cases} \mathbf{b}^\top \mathbf{z} & \text{if } \mathbf{B}^\top \mathbf{z} = \boldsymbol{\nu},\ \mathbf{z} \geq 0 \\ \infty & \text{otherwise} \end{cases}$

similar to model (1.8). This problem is now again a standard robust optimization problem. We may, without loss of generality, consider the uncertainty constraint-wise, see Ben-Tal et al. (2009), in order to derive the tractable affinely adjustable robust counterpart (AARC) for each constraint i:

$$\forall \boldsymbol{\zeta} \in \mathcal{Z}: \quad (\mathbf{a}_i + \mathbf{A}_i\boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}_i^\top(\mathbf{u} + \mathbf{V}\boldsymbol{\zeta}) \leq d_i,$$

which is equivalent to

$$\mathbf{a}_i{}^\top \mathbf{x} + \mathbf{b}_i^\top \mathbf{u} + \max_{\boldsymbol{\zeta}\in\mathcal{Z}} \left\{(\mathbf{A}_i^\top \mathbf{x} + \mathbf{V}^\top \mathbf{b}_i)^\top \boldsymbol{\zeta}\right\} \leq d_i.$$

or

$$\mathbf{a}_i{}^\top \mathbf{x} + \mathbf{b}_i^\top \mathbf{u} + \delta^*(\mathbf{A}_i^\top \mathbf{x} + \mathbf{V}^\top \mathbf{b}_i \mid \mathcal{Z}) \leq d_i.$$

where $\delta^*(\boldsymbol{\nu} \mid \mathcal{Z}) = \max_{\boldsymbol{\zeta}\in\mathcal{Z}}\{\boldsymbol{\zeta}^\top\boldsymbol{\nu}\}$ is the so-called *support function* of the set $\mathcal{Z}$. The notation δ^* is the conjugate function of the indicator function

$$\delta(\boldsymbol{\zeta} \mid \mathcal{Z}) = \begin{cases} 0 & \text{if } \boldsymbol{\zeta} \in \mathcal{Z} \\ \infty & \text{otherwise.} \end{cases}$$

For many different closed convex sets $\mathcal{Z}$ the support function can be explicitly constructed. Some examples are given in Table 6.1 and many more can be found in Ben-Tal et al. (2015).

6.3 The new adjustable robust model based on inexact data

This section introduces our model that extends the ARC model to the case where revealed data is inexact. We stress that the models described here are more general and not limited to production-inventory problems. They could be used for any ARO problem within operations management where the revealed data is inexact.

The ARO model with decision rules based on exact data assumes that there is one moment in time where the data $\boldsymbol{\zeta} \in \mathcal{Z}$, used to decide upon the variable $\mathbf{y}$, is known exactly. However, in many practical applications only an estimate $\widehat{\boldsymbol{\zeta}} \in \mathcal{Z}$ of the true value $\boldsymbol{\zeta}$ can be obtained. In that case we have *inexact* data and $\widehat{\boldsymbol{\zeta}}$ is not exactly equal to $\boldsymbol{\zeta}$, but we may assume that the estimation error $\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}$ resides in another closed convex set $\widehat{\mathcal{Z}}$, which we call the *estimation uncertainty*. We also denote this as $\boldsymbol{\zeta} \in \{\widehat{\boldsymbol{\zeta}}\} + \widehat{\mathcal{Z}}$, the Minkowski sum of a singleton and a set. Note that estimation errors of different components of $\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}$ can be correlated. The decision rule for the wait-and-see variable is only allowed to use the estimate $\widehat{\boldsymbol{\zeta}}$ (and not the unobserved $\boldsymbol{\zeta}$):

$$\mathbf{y} = \mathbf{u} + \mathbf{V}\widehat{\boldsymbol{\zeta}},$$

where (here-and-now) decision variables $\mathbf{u}$ and $\mathbf{V}$ determine the affine dependence of $\mathbf{y}$ on estimate $\widehat{\boldsymbol{\zeta}}$. We call the robust counterpart in this new setting the (affine) adjustable robust counterpart with decision rules *based on inexact data*, or ARCID:

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u},\mathbf{V}} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \forall(\boldsymbol{\zeta},\widehat{\boldsymbol{\zeta}}) \in \mathcal{U}: \quad (\mathbf{a}_i + \mathbf{A}_i\boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}_i^\top(\mathbf{u} + \mathbf{V}\widehat{\boldsymbol{\zeta}}) \le d_i \qquad \forall i = 1,\dots,J, \end{aligned} \tag{ARCID}$$

where

$$\mathcal{U} = \{(\boldsymbol{\zeta},\widehat{\boldsymbol{\zeta}}): \ \boldsymbol{\zeta},\widehat{\boldsymbol{\zeta}} \in \mathcal{Z}, \ (\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}) \in \widehat{\mathcal{Z}}\} \tag{6.1}$$

provides us with a new uncertainty set that describes in a general way the uncertain parameter $\boldsymbol{\zeta}$, its estimate $\widehat{\boldsymbol{\zeta}}$ and the relation between these two uncertain vectors. Note that the set $\mathcal{U}$ is closed and convex whenever the sets $\mathcal{Z}$ and $\widehat{\mathcal{Z}}$ are closed and convex. The relation between the RC, the new ARCID and the classical ARC uncertainty sets in terms of the inexactness in the revealed data, is depicted in Figure 6.1. In the RC none of the revealed information is used, so it assumes that the parameter can still take any value in the uncertainty set when deciding upon $\mathbf{y}$. The ARCID uses the revealed information and takes into account that the data used in the decision rule is inexact and therefore is still uncertain to some extent. The ARC model also uses the revealed information, but does however assume that these data are exact. The implications of this assumption, when in reality the observed information is inexact, shall become clear in our numerical example in Section 6.4. Note that in the uncertainty described in (6.1) both the true parameter and its estimate are in the set $\mathcal{Z}$. Another modelling choice could be to leave out any further condition on the estimate and just have $(\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}) \in \widehat{\mathcal{Z}}$. Omitting this condition $\widehat{\boldsymbol{\zeta}} \in \mathcal{Z}$, however, leads to an increase of the size of the uncertainty set for the estimate. In that case, the decision rule should be valid on a larger uncertainty set which might lead to more conservative solutions. Furthermore, some values for the estimates can be naturally omitted. For example, demand is nonnegative and any negative estimates can be rounded up to zero. As in the previous ARC setting we consider, without

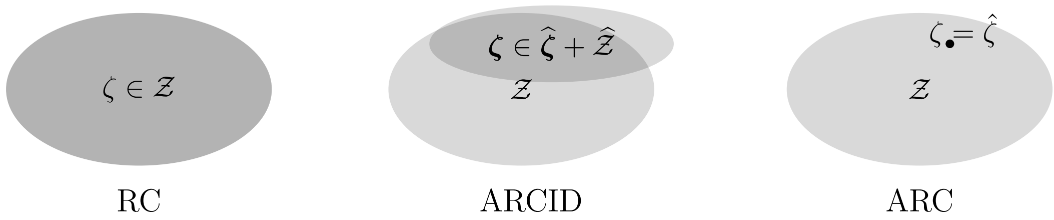


Figure 6.1 – Comparison between uncertainty of the revealed information in the RC, ARCID and ARC concepts.

loss of generality, constraint-wise uncertainty. Hence, we only have to determine the tractable formulation of the i-th constraint

$$\forall(\boldsymbol{\zeta},\widehat{\boldsymbol{\zeta}}) \in \mathcal{U}: \quad (\mathbf{a}_i + \mathbf{A}_i\boldsymbol{\zeta})^\top \mathbf{x} + \mathbf{b}_i^\top(\mathbf{u} + \mathbf{V}\hat{\boldsymbol{\zeta}}) \leq d_i, \tag{6.2}$$

which follows from the next theorem.

Theorem 6.1 *Let $\mathcal{U}$ be a closed set with nonempty relative interior as given in* (6.1). *Then $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{V})$ satisfies constraint* (6.2) *if and only if there exists a $\boldsymbol{w}_i \in \mathbb{R}^L$ that satisfies*

$$\boldsymbol{a_i}^\top\boldsymbol{x} + \boldsymbol{b}_i^\top\boldsymbol{u} + \delta^*(\boldsymbol{A}_i^\top\boldsymbol{x} - \boldsymbol{w}_i \mid \mathcal{Z}) + \delta^*(\boldsymbol{V}^\top\boldsymbol{b}_i + \boldsymbol{w}_i \mid \mathcal{Z}) + \delta^*(\boldsymbol{w}_i \mid \widehat{\mathcal{Z}}) \leq d_i.$$

Proof. We can replace the semi-infinite constraint (6.2) by constraints involving maximization over the uncertainty and obtain the following constraint:

$$\mathbf{a}_i{}^\top\mathbf{x} + \mathbf{b}_i^\top\mathbf{u} + \max_{(\zeta,\widehat{\zeta})\in\mathcal{U}} \left\{ \begin{pmatrix} \mathbf{A}_i^\top\mathbf{x} \\ \mathbf{V}^\top\mathbf{b}_i \end{pmatrix}^\top \begin{pmatrix} \zeta \\ \widehat{\zeta} \end{pmatrix} \right\} \leq d_i,$$

or, by using the definition of support functions,

$$\mathbf{a}_i{}^\top\mathbf{x} + \mathbf{b}_i^\top\mathbf{u} + \delta^*\left(\begin{pmatrix} \mathbf{A}_i^\top\mathbf{x} \\ \mathbf{V}^\top\mathbf{b}_i \end{pmatrix} \mid \mathcal{U} \right) \leq d_i. \tag{6.3}$$

Hence, all we need to do is to find an expression for the support function of $\mathcal{U}$. To do so, note that for the indicator function we have $\delta\left(\begin{pmatrix} \zeta \\ \widehat{\zeta} \end{pmatrix} \mid \mathcal{U}\right) = \delta(\boldsymbol{\zeta} \mid \mathcal{Z}) + \delta(\widehat{\boldsymbol{\zeta}} \mid \mathcal{Z}) + \delta\left((\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}) \mid \widehat{\mathcal{Z}}\right)$. If we define the function $h(\boldsymbol{\zeta},\widehat{\boldsymbol{\zeta}}) = \delta\left((\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}) \mid \widehat{\mathcal{Z}}\right)$, then by using the definition of conjugate functions as in Rockafellar (1970), we can obtain its conjugate function:

$$h^*(\boldsymbol{w}_i, \widetilde{\boldsymbol{w}}_i) = \begin{cases} \delta^*\left(\boldsymbol{w}_i \mid \widehat{\mathcal{Z}}\right) & \text{if } \boldsymbol{w}_i + \widetilde{\boldsymbol{w}}_i = 0 \\ \infty & \text{otherwise.} \end{cases}$$

Using this conjugate function, and the fact that $\mathcal{U}$ has nonempty relative interior, we can now find the expression for the support function in (6.3) using the conjugate of a sum of functions (see Rockafellar (1970, Chapter 16)):

$$\delta^*\left(\begin{pmatrix}\mathbf{A}_i^\top \mathbf{x}\\ \mathbf{V}^\top \mathbf{b}_i\end{pmatrix} \mid \mathcal{U}\right) = \min_{\boldsymbol{w}_i,\widetilde{\boldsymbol{w}}_i,\boldsymbol{z}_i,\widetilde{\boldsymbol{z}}_i}\Big\{\delta^*(\boldsymbol{z}_i \mid \mathcal{Z}) + \delta^*(\widetilde{\boldsymbol{z}}_i \mid \mathcal{Z}) + h^*(\boldsymbol{w}_i,\widetilde{\boldsymbol{w}}_i) \mid \boldsymbol{w}_i + \boldsymbol{z}_i = \mathbf{A}_i^\top \mathbf{x},\ \widetilde{\boldsymbol{w}}_i + \widetilde{\boldsymbol{z}}_i = \mathbf{V}^\top \mathbf{b}_i\Big\}$$

$$= \min_{\boldsymbol{w}_i,\widetilde{\boldsymbol{w}}_i,\boldsymbol{z}_i,\widetilde{\boldsymbol{z}}_i}\Big\{\delta^*(\boldsymbol{z}_i|\mathcal{Z}) + \delta^*(\widetilde{\boldsymbol{z}}_i|\mathcal{Z}) + \delta^*(\boldsymbol{w}_i \mid \widehat{\mathcal{Z}}) \mid \boldsymbol{w}_i + \boldsymbol{z}_i = \mathbf{A}_i^\top \mathbf{x},\ \widetilde{\boldsymbol{w}}_i + \widetilde{\boldsymbol{z}}_i = \mathbf{V}^\top \mathbf{b}_i,\ \boldsymbol{w}_i + \widetilde{\boldsymbol{w}}_i = 0\Big\}.$$

Substituting this result in (6.3) yields that (6.2) is feasible if and only if there exist $\boldsymbol{w}_i, \widetilde{\boldsymbol{w}}_i, \boldsymbol{z}_i, \widetilde{\boldsymbol{z}}_i \in \mathbb{R}^L$ that satisfy

$$\begin{cases} \mathbf{a}_i{}^\top \mathbf{x} + \mathbf{b}_i^\top \mathbf{u} + \delta^*(\boldsymbol{z}_i \mid \mathcal{Z}) + \delta^*(\widetilde{\boldsymbol{z}}_i \mid \mathcal{Z}) + \delta^*(\boldsymbol{w}_i \mid \widehat{\mathcal{Z}}) \le d_i \\ \boldsymbol{w}_i + \boldsymbol{z}_i = \mathbf{A}_i^\top x \\ \widetilde{\boldsymbol{w}}_i + \widetilde{\boldsymbol{z}}_i = \mathbf{V}^\top \mathbf{b}_i \\ \boldsymbol{w}_i + \widetilde{\boldsymbol{w}}_i = 0. \end{cases}$$

The result then follows by elimination of the variables $\widetilde{\boldsymbol{w}}_i$, $\boldsymbol{z}_i$ and $\widetilde{\boldsymbol{z}}_i$. ■

The two assumptions on the uncertainty set (closedness and nonempty relative interior of $\mathcal{U}$) used in Theorem 6.1 are satisfied for all closed sets $\mathcal{Z}$ and $\widehat{\mathcal{Z}}$ with nonempty relative interior and 0 being an element of the relative interior of $\widehat{\mathcal{Z}}$. A few common choices for uncertainty sets, that satisfy these conditions, have been given in Table 6.1. Below we give two examples of constraints with different choices for the estimation uncertainty. In the first example (Box-Box) we have both box uncertainty for the parameter $\boldsymbol{\zeta}$ and a box for the estimation error (independent estimation errors). In the second example (Box-Ball) the estimation errors reside in a ball.

Example 6.1 (Box-Box) *If $\mathcal{Z} = \{\boldsymbol{\zeta} : ||\boldsymbol{\zeta}||_\infty \le \theta\}$ and $\widehat{\mathcal{Z}} = \{\boldsymbol{\xi} : ||\boldsymbol{\xi}||_\infty \le \rho\}$ for some scalar uncertainty levels $\theta, \rho \ge 0$ then, according to Theorem 6.1, $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{V})$ satisfies constraint* (6.2) *if and only if there exists a $\boldsymbol{w}_i \in \mathbb{R}^L$ such that*

$$\boldsymbol{a}_i{}^\top \boldsymbol{x} + \boldsymbol{b}_i^\top \boldsymbol{u} + \theta||\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{w}_i||_1 + \theta||\,\boldsymbol{V}^\top \boldsymbol{b}_i + \boldsymbol{w}_i||_1 + \rho||\boldsymbol{w}_i||_1 \le d_i,$$

where the expressions for the support functions with these choices for the uncertainty sets are found using Table 6.1. This constraint can be represented by a set of linear constraints.

Example 6.2 (Box-Ball) *If $\mathcal{Z} = \{\boldsymbol{\zeta} : ||\boldsymbol{\zeta}||_\infty \le \theta\}$ and $\widehat{\mathcal{Z}} = \{\boldsymbol{\xi} : ||\boldsymbol{\xi}||_2 \le \rho\}$ for some scalar uncertainty levels $\theta, \rho \ge 0$ then, according to Theorem 6.1, $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{V})$ satisfies constraint* (6.2) *if and only if there exists a $\boldsymbol{w}_i \in \mathbb{R}^L$ such that*

$$\boldsymbol{a}_i{}^\top \boldsymbol{x} + \boldsymbol{b}_i^\top \boldsymbol{u} + \theta||\boldsymbol{A}_i^\top \boldsymbol{x} - \boldsymbol{w}_i||_1 + \theta||\,\boldsymbol{V}^\top \boldsymbol{b}_i + \boldsymbol{w}_i||_1 + \rho||\boldsymbol{w}_i||_2 \le d_i,$$

where the expressions for the support functions with these choices for the uncertainty sets are again found using Table 6.1. This constraint can be represented by a set of linear constraints and a conic quadratic constraint.

Theorem 6.1 can also be used to argue that the new ARCID model bridges the gap between models that do not use information at all in the second stage (RC) and those that rely on fully accurate revealed information in the decision rules (ARC). Namely, if the estimation uncertainty is large (i.e. $\widehat{\mathcal{Z}}$ is large), then there is no value in the revealed inexact data. In that case the optimal value of the nonadjustable version is equal to the optimal value of (ARCID). More formally, consider the situation where there exists a realization $\bar{\boldsymbol{\zeta}} \in \mathcal{Z}$ such that $\mathcal{Z} \subset \bar{\boldsymbol{\zeta}} + \widehat{\mathcal{Z}}$. Then, if (ARCID) is feasible, it follows directly that there must also exist a decision rule with $\mathbf{V} = 0$, i.e., a nonadjustable decision. For Example 6.1 and 6.2 we have that the ARCID model is equivalent to the nonadjustable model when $\rho \geq \theta$ for the first example (Box-Box) and $\rho \geq \sqrt{L}\theta$ for the second example (Box-Ball). In case there is no estimation error ($\widehat{\mathcal{Z}} = \{0\}$), the ARC and the ARCID are equivalent in the sense that they have the same feasible region and the same optimal objective value.

So far, we have focussed on the two period case for illustrative purposes. However, often we have multiple periods $1, 2, \ldots, T$, in which we consecutively have to make decisions $\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^T$. In period t we can make decisions based on estimates available in that period: estimate $\widehat{\boldsymbol{\zeta}}^t$. So, we have in period t a linear decision rule $\mathbf{y}^t = \mathbf{u}^t + \mathbf{V}^t \boldsymbol{R}^t \widehat{\boldsymbol{\zeta}}^t$, with variables $\mathbf{u}^t \in \mathbb{R}^m$ and $\mathbf{V}^t \in \mathbb{R}^{m \times L}$. The matrix $\boldsymbol{R}^t \in \mathbb{R}^{L \times L}$ is the (fixed) diagonal *information matrix* with entries 0 everywhere but on the diagonal. The entries on the diagonal are either 0 (if no data is revealed) or 1 if the estimate is available at time t. For the standard case in the literature, with decision rules based on exact data $\boldsymbol{\zeta}$, we have

$$\forall \boldsymbol{\zeta} \in \mathcal{Z}: \quad (\mathbf{a}_i + \mathbf{A}_i \boldsymbol{\zeta})^\top \mathbf{x} + \sum_{t=1}^{T} (\boldsymbol{b}_i^t)^\top (\mathbf{u}^t + \mathbf{V}^t \boldsymbol{R}^t \boldsymbol{\zeta}) \leq d_i.$$

Note that the true parameter has the same (unknown) value over all periods $t = 1, \ldots, T$, only the information matrix might change. If we now take into account the inexact nature of our estimates, i.e., basing decision in period t on the observed estimate $\widehat{\boldsymbol{\zeta}}^t$, this constraint becomes

$$\forall (\boldsymbol{\zeta}, \widehat{\boldsymbol{\zeta}}^1, \widehat{\boldsymbol{\zeta}}^2, \ldots, \widehat{\boldsymbol{\zeta}}^T) \in \mathcal{U}_T: \quad (\mathbf{a}_i + \mathbf{A}_i \boldsymbol{\zeta})^\top \mathbf{x} + \sum_{t=1}^{T} (\boldsymbol{b}_i^t)^\top (\mathbf{u}^t + \mathbf{V}^t \boldsymbol{R}^t \widehat{\boldsymbol{\zeta}}^t) \leq d_i, \tag{6.4}$$

which is the multistage equivalent of constraint (6.2) with uncertainty set

$$\mathcal{U}_T = \left\{ (\boldsymbol{\zeta}, \widehat{\boldsymbol{\zeta}}^1, \widehat{\boldsymbol{\zeta}}^2, \ldots, \widehat{\boldsymbol{\zeta}}^T) \ : \ \boldsymbol{\zeta}, \widehat{\boldsymbol{\zeta}}^1, \widehat{\boldsymbol{\zeta}}^2, \ldots, \widehat{\boldsymbol{\zeta}}^T \in \mathcal{Z}, \ (\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}^t) \in \widehat{\mathcal{Z}}_t \quad \forall t \right\}, \tag{6.5}$$

where $\widehat{\mathcal{Z}}_t$ describes the estimation uncertainty for $\widehat{\boldsymbol{\zeta}}^t$, which is the estimate of $\boldsymbol{\zeta}$ in period t. We can readily extend Theorem 6.1 to these types of constraints in multistage problems. The proof is similar to the proof for the two period case and can be found in Appendix 6.A.

Theorem 6.2 *Let $\mathcal{Z}$, $\widehat{\mathcal{Z}}_1, \ldots, \widehat{\mathcal{Z}}_T$ be closed sets with nonempty interior as given in* (6.5)*. Then* $(\boldsymbol{x}, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^T, \boldsymbol{V}^1, \ldots, \boldsymbol{V}^T)$ *satisfies* (6.4) *if and only if there exist* $\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iT} \in \mathbb{R}^L$ *that satisfy*

$$\boldsymbol{a}_i^\top \boldsymbol{x} + \sum_{t=1}^{T} (\boldsymbol{b}_i^t)^\top \boldsymbol{u}^t + \delta^* \left(\boldsymbol{A}_i^\top \boldsymbol{x} - \sum_{t=1}^{T} \boldsymbol{w}_{it} \mid \mathcal{Z} \right) + \ldots$$

$$\sum_{t=1}^{T} \delta^* \left((\boldsymbol{V}^t \boldsymbol{R}^t)^\top \boldsymbol{b}_i^t + \boldsymbol{w}_{it} \mid \mathcal{Z} \right) + \sum_{t=1}^{T} \delta^* \left(\boldsymbol{w}_{it} \mid \widehat{\mathcal{Z}}_t \right) \leq d_i.$$

In Theorem 6.1 we only consider constraints that are linear. This theorem can be readily extended to the case where the constraint is convex (but not necessarily linear) in the here-and-now variables $\mathbf{x}$. To do so, we can use Fenchel duality as has been done for nonadjustable robust models in Ben-Tal et al. (2015).

The construction of the standard uncertainty set and the estimation uncertainty set can be done in different ways. Our model based on inexact revealed data has additional uncertainty in the estimates described by the uncertainty sets $\hat{\mathcal{Z}}$ or $\widehat{\mathcal{Z}}_1, \ldots, \widehat{\mathcal{Z}}_T$ in the multiperiod case. We have to construct another uncertainty set that captures all estimation errors for which we want to be protected in our future planning periods. For constructing the estimation uncertainty set we can use the same techniques as for the static case (see e.g. Bertsimas et al. (2017a)). We can for instance use historical data on the errors, $\boldsymbol{\zeta} - \widehat{\boldsymbol{\zeta}}^t$, obtained from previous planning horizons. If there is insufficient historical data, one can still define uncertainty sets with realistic a priori reasoning. In retail stores, and especially with the growing share of online retail, customers often return a product if it does not meet their requirements. Sales figures then give an indication of the total demand, but it is known that in each period between, for example, 5% and 10% of all products are returned. The bandwidth of this percentage can then be used to construct the estimation uncertainty around the demand estimate obtained via sales figures. Another situation of estimation uncertainty arises when the demand estimate is obtained via accumulation of (correlated) demand from different stores. If we know that different stores need different amounts of time to come up with accurate data (e.g., sales reports), then there is still some uncertainty on the total demand if, for example, only 9 out of 10 stores have reported their sales. In both of these described situations more information will be revealed in later periods and estimates are likely to become more accurate over time. An example of this type of uncertainty set where estimates become more accurate over time is used in the production-inventory problem in the next section.

6.4 Production-inventory problem

In this section we apply the ARCID approach to the production-inventory problem that was introduced in Ben-Tal et al. (2004), the seminal paper on adjustable robust optimization.

6.4.1 The nominal model

We consider a single product inventory system, which is comprised of a warehouse and I factories. A planning horizon of T periods is used. In the model we use the following parameters and variables, using the same notation as in Ben-Tal et al. (2004):

Parameters

- d_t Demand for the product in period t;
- $P_i(t)$ Production capacity of factory i in period t;
- $c_i(t)$ Costs of producing one product unit at factory i in period t;
- $V_{\min}$ Minimal allowed level of inventory at the warehouse;
- $V_{\max}$ Storage capacity of the warehouse;
- Q_i Cumulative production capacity of the i-th factory throughout the planning horizon.

Variables

- $p_i(t)$ The amount of the product to be produced in factory i in period t;
- $v(t)$ Inventory level at the beginning of period t ($v(1)$ is given).

We try to minimize the total production costs over all factories and the whole planning horizon. The restriction is that all demand in period t must be satisfied by units in stock in the warehouse or by the production in period t. If all the demand, and all other parameters, are certain in all periods $1, \ldots, T$, then the problem is modeled by the following linear program (Ben-Tal et al. 2004, Section 5) Section 5:

$$\begin{aligned}
\min_{p_i(t), v(t), F} \quad & F \\
\text{s.t.} \quad & \sum_{t=1}^{T} \sum_{i=1}^{I} c_i(t) p_i(t) \leq F \\
& 0 \leq p_i(t) \leq P_i(t), \qquad \forall i = 1, \ldots, I, \forall t = 1, \ldots, T \\
& \sum_{t=1}^{T} p_i(t) \leq Q_i, \qquad \forall i = 1, \ldots, I \\
& v(t+1) = v(t) + \sum_{i=1}^{I} p_i(t) - d_t, \qquad \forall t = 1, \ldots, T \\
& V_{\min} \leq v(t) \leq V_{\max}, \qquad \forall t = 2, \ldots, T+1.
\end{aligned} \tag{P:Nominal}$$

6.4.2 The affinely adjustable robust model based on inexact data

We assume that we can make decisions based on estimates of the realized demand scenario $\mathbf{d} = (d_1, \ldots, d_T)$. We should specify our production policies for the factories before the planning periods starts, at period 0. When we specify these policies, we only know that demand in consecutive periods are independent and reside in a certain box region,

$$d_t \in \mathcal{Z}_t = [d_t^* - \theta d_t^*, d_t^* + \theta d_t^*], \qquad t = 1, \ldots, T, \tag{6.6}$$

with given $0 < \theta \leq 1$, the level of uncertainty, and nominal demand d_t^* in period t. So far the model is exactly the same as in Ben-Tal et al. (2004) if we assume that we can estimate the demand d_t exactly in periods $r \in I_t$, where I_t is a given subset of $\{1, \ldots, T\}$. In Ben-Tal et al. (2004) different sets for I_t are used:

- $I_t = \{1, \ldots, t\}$, the information basis where demand from the past and the present is known exactly, for the future no extra information is known;
- $I_t = \{1, \ldots, t-1\}$, the information basis where all demand from the past is known exactly, there is no information about the present;
- $I_t = \{1, \ldots, t-4\}$, the information about the past is received with a four day delay. For other periods in the past ($t-3, t-2$ and $t-1$) there is no extra information at all.

Now we assume the decisions in period t are based on estimates $\widehat{d}_{r,t}$, made in period t, for the actual demand d_r in the period $r \in \{1, \ldots, t\}$. We assume that these estimates can, in principle, take any value that the demand d_r can take, so $\widehat{d}_{r,t} \in \mathcal{Z}_r$ and that the estimation error $\widehat{d}_{r,t} - d_r$ lies in a box region:

$$\widehat{d}_{r,t} - d_r \in \widehat{\mathcal{Z}}_{r,t} = [-\rho_{r,t}\theta d_t^*, \rho_{r,t}\theta d_t^*], \tag{6.7}$$

where the parameter $\rho_{r,t}$ indicates the fraction of initial uncertainty level θ for the estimate $\widehat{d}_{r,t}$.

Note that if we have exact information for periods in the information basis, i.e., $\widehat{d}_{r,t} = d_t$ for all $r \in I_t$ and no extra information (besides $\widehat{d}_{r,t} \in \mathcal{Z}_t$) for all periods outside the information basis, then we end up in the case of exact revealed information as considered by Ben-Tal et al. (2004). This situation can be can be modeled as a special case of our model by using the following values for $\rho_{r,t}$

$$\rho_{r,t} = \begin{cases} 0 & \text{if } r \in I_t \\ 1 & \text{otherwise,} \end{cases}$$

which means that the estimation error equals zero for estimates on demand in periods that lie in the information set and it is θ (so very large) for periods outside this information basis.

The general situation with inexact data lies in between the two extreme scenarios where one either knows the demand exact, or not at all. For this we specify the information set in a more general way:

$$\hat{I}_t := \{r : \rho_{r,t} < 1\}.$$

This definition of $\hat{I}_t$ is indeed a more general description. For large estimation errors ($\rho_{r,t} \geq 1$) we could just as well decide on the variables beforehand, i.e., we have no extra useful information on the actual realizations compared to the information at time $t = 0$. We can therefore safely exclude all periods where the estimates are too noisy (the periods for which $r \notin \hat{I}_t$). Since we apply the ARCID method based on inexact data, we take affine decision rules based on *inexact* estimates:

$$p_i(t) = \pi^0_{i,t} + \sum_{r \in \hat{I}_t} \pi^r_{i,t} \widehat{d}_{r,t}, \tag{6.8}$$

where the coefficients $\pi^r_{i,t}$ are the new nonadjustable variables in the model. For notational convenience we write the vector $\widehat{\mathbf{d}}_t$ as the vector containing all the estimates $\widehat{d}_{r,t}$ for all $r \in \hat{I}_t$, $t = 1, \dots, T$. The uncertainty set can now be written as:

$$\mathcal{U} := \left\{ (\mathbf{d}, \widehat{\mathbf{d}}_1, \dots, \widehat{\mathbf{d}}_T) : d_r, \widehat{d}_{r,t} \in \mathcal{Z}_r, (\widehat{d}_{r,t} - d_r) \in \widehat{\mathcal{Z}}_{r,t}, \quad \forall r \in \hat{I}_t, \quad \forall t \right\},$$

with $\mathcal{Z}_t$ and $\widehat{\mathcal{Z}}_{r,t}$ as specified in respectively (6.6) and (6.7). The linear problem (P:Nominal) becomes (after elimination of the v-variables) a semi-infinite LP if we use linear decision rule (6.8):

$$\begin{aligned}
&\min_{\pi,F} \quad F \\
&\text{s.t.} \quad \forall (\mathbf{d}, \widehat{\mathbf{d}}_1, \dots, \widehat{\mathbf{d}}_T) \in \mathcal{U} : \\
&\begin{cases}
\sum_{t=1}^{T} \sum_{i=1}^{I} c_i(t) \left(\pi^0_{i,t} + \sum_{r \in \hat{I}_t} \pi^r_{i,t} \widehat{d}_{r,t} \right) \leq F \\
0 \leq \pi^0_{i,t} + \sum_{r \in \hat{I}_t} \pi^r_{i,t} \widehat{d}_{r,t} \leq P_i(t), \qquad \forall i, t \\
\sum_{t=1}^{T} \left(\pi^0_{i,t} + \sum_{r \in \hat{I}_t} \pi^r_{i,t} \widehat{d}_{r,t} \right) \leq Q_i, \qquad \forall i \\
V_{\min} \leq v(1) + \sum_{s=1}^{t} \sum_{i=1}^{I} \left(\pi^0_{i,s} + \sum_{r \in \hat{I}_t} \pi^r_{i,s} \widehat{d}_{r,s} \right) - \sum_{s=1}^{t} d_s \leq V_{\max}, \ \forall t.
\end{cases}
\end{aligned} \tag{P:ARCID}$$

The resulting tractable robust counterpart can be found using Theorem 6.2 and is given in Appendix 6.B.

6.4.3 Data set from Ben-Tal et al. (2004)

We take the same data set as in the illustrative example by Ben-Tal et al. (2004, p.370-371):

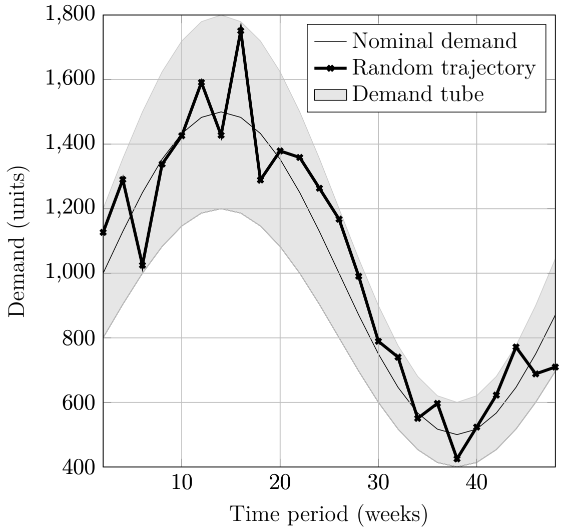


Figure 6.2 – Demand.

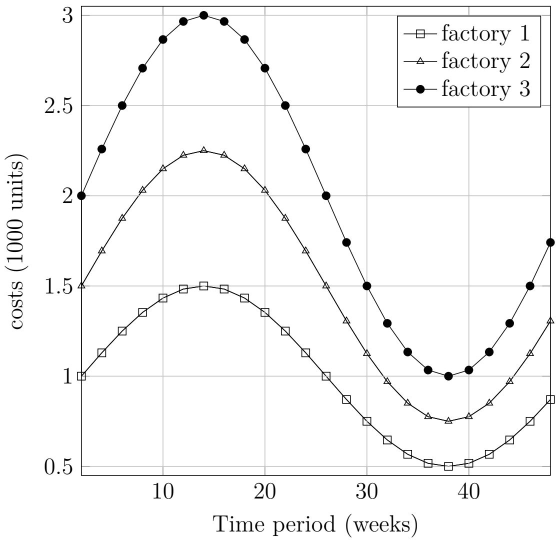


Figure 6.3 – Costs.

"There are $I = 3$ factories producing a seasonal product, and one warehouse. The decisions concerning production are made every two weeks, and we are planning production for 48 weeks, thus the time horizon is $T = 24$ periods. The nominal demand d^* is seasonal, reaching its maximum in winter, specifically,

$$d_t^* = 1000\left(1 + \tfrac{1}{2}\sin\left(\tfrac{\pi(t-1)}{12}\right)\right), \quad t = 1, \ldots, 24.$$

We assume that the uncertainty level θ is 20%, i.e., $d_t \in [0.8d_t^*, 1.2d_t^*]$, as shown on Figure 6.2. The production costs per unit of the product depend on the factory and on time and follow the same seasonal pattern as the demand, i.e., rise in winter and fall in summer. The production costs for a factory i at a period t is given by:

$$c_i(t) = \alpha_i\left(1 + \tfrac{1}{2}\sin\left(\tfrac{\pi(t-1)}{12}\right)\right), \quad t = 1, \ldots, 24.$$
$$\alpha_1 = 1$$
$$\alpha_2 = 1.5$$
$$\alpha_3 = 2$$

The maximal production capacity of each one of the factories at each two-weeks period is $P_i(t) = 567$ units, and the integral production capacity of each one of the factories for a year is $Q_i = 13600$. The inventory at the warehouse should be no less than 500 units, and cannot exceed 2000 units."

The initial inventory level $v(1)$ was not stated in Ben-Tal et al. (2004), but this value is equal to the lower bound of the inventory level at the warehouse, namely 500. Note that the initial inventory level could also be chosen uncertain if the initial state is unknown. For new products, where no past demand has occurred, it is realistic to assume no uncertainty on the stock as the inventory level is set by the manager itself. Here we also assume that the initial inventory level is known, as in Ben-Tal et al. (2004).

6.5 Numerical results

Ben-Tal et al. (2004) conduct two series of experiments based on the data given in Section 6.4.3. In the first series of experiments they modify the parameter θ to analyze the influence of demand uncertainty on the total production cost. In the second series of experiments they change the information basis I_t, the (exact) information that is used in the decision rule. Note that Ben-Tal et al. (2004) deal with the case where in period t all demand from the periods in the information set I_t is known *exactly*. For instance, if the information set is equal to $I_t = \{1, \ldots, t-1\}$, then in period t we can base our production decision rule on the exact values of the demand realizations in periods $1, \ldots, t-1$, and use no information on the demand in periods after $t-1$. We extend these experiments to include inexact data in some periods to show the benefits of the ARCID model over the ARC model.

Just as in Ben-Tal et al. (2004), we test the management policies by simulating 100 demand trajectories, $d = (d_1, \ldots, d_T)$. For every simulation the demand trajectory is randomly generated with d_t uniformly distributed in $[(1-\theta)d_t^*, (1+\theta)d_t^*]$, where 20% ($\theta = 0.2$) is the chosen uncertainty level. The uncertainty level of the demand is set to 20% in all experiments, as this seems to be the most restrictive level of uncertainty and is the same level that has been used by Ben-Tal et al. (2004). For higher uncertainty levels like 30%, even the model without uncertainty (P:Nominal) is no longer feasible for the maximal demand pattern with $d_t = (1+\theta)d_t^*$ (without uncertainty) because of the bounds on production imposed by $P_i(t)$ and Q_i. In line with the experiments performed by Ben-Tal et al. (2004), we compute the *average* costs for our solutions by assuming an uniform distribution for the estimated demand. In Ben-Tal et al. (2004) they have used 100 simulated demand trajectories to approximate the mean costs. However, since the costs are linear in the estimated demand parameter, this can be found by substituting the expected (nominal) demand in the objective function. All solutions are obtained by the commercial solver Gurobi Optimization (2015) programmed in the YALMIP language (Löfberg 2004) in MATLAB.

6.5.1 Experiments with decision rules using inexact data on demand

Similar to Ben-Tal et al. (2004), we saved the demand trajectories to compute the so-called costs of the ideal setting, the utopian world where the entire demand trajectory is known beforehand. The ideal setting is used to benchmark the performance of the ARCID solution. In the ideal setting one sets the policy only for one sample demand

Table 6.2 – The influence of the estimation errors on the mean costs and worst case costs (WC) in the ARCID model.
(The dashes represent estimation errors of 100%)

	Demand estimation error $\rho_{r,t}$ (in %)										Costs	
Case	$\rho_{1,t},\ldots,\rho_{t-9,t}$	$\rho_{t-8,t}$	$\rho_{t-7,t}$	$\rho_{t-6,t}$	$\rho_{t-5,t}$	$\rho_{t-4,t}$	$\rho_{t-3,t}$	$\rho_{t-2,t}$	$\rho_{t-1,t}$	$\rho_{t,t}$	Mean	WC
1	0	0	0	0	0	0	0	0	0	10	35,167	44,268
2	0	0	0	0	0	0	0	0	0	20	35,077	44,273
3	0	0	0	0	0	0	0	0	20	-	35,740	44,582
4	0	0	0	0	0	0	0	0	-	-	35,740	44,582
5	0	0	0	0	0	0	1	5	10	-	36,882	44,883
6	0	0	0	5	5	5	10	10	10	-	36,867	45,326

realization, so the solution does not have to be feasible for all possible demand trajectories. Hence, the costs in the ideal setting are obviously a lower bound of the costs for the ARCID solutions. For the ideal setting the worst case is the demand trajectory with the highest demand: $d_t = (1+\theta)d_t^*$ for all t. The worst case costs in the ideal setting can be easily solved and turns out to be 44,199. The mean costs in the ideal case are approximated by averaging the ideal costs for the 100 simulated demand trajectories and equals 33,729.

In our model, the demand from the past periods is not known exactly, but we assume to have inexact estimates for some past and present periods. Several cases are investigated, for instance those where the delay for receiving the exact demand information is even more than 2 periods, i.e., the *exact* demand is known after 3, 4 or more periods. These cases are infeasible in the ARC model, see Ben-Tal et al. (2004).

In the experiments, the influence of the estimation error $\rho_{r,t}$ on the total production costs is tested. An estimation error of 0% for the demand in period $t-1$ means that $\rho_{t-1,t} = 0$ (exact information). An estimation uncertainty of 10% for the demand in period $t-4$ means that $\rho_{t-4,t} = 0.1$ and so forth. We have considered various estimation uncertainties for the estimates on past realizations, as depicted in Table 6.2. Note that in all cases the estimates become more accurate over time. In other words, the estimation error decreases over time: $\rho_{t-r,t} \leq \rho_{t-s,t}$ for all $r \leq s$ and all periods t. In Table 6.2 one notices this by seeing that the values for the estimation errors are decreasing right-to-left. Therefore, estimates on demand values from longer ago in the past are more accurate than estimates on recent demand realizations.

The cases in Table 6.2 can be explained as follows:

- For Cases 1 and 2 we assume that all demand from the past is known exactly. For the present period we have a good estimate on the demand that gives extra information compared to the information known at the start of the planning period ($t = 0$).

- The Cases 3-6 assume to have no additional knowledge about the present. Furthermore, the exact demand from previous periods is received with a certain delay, but there are already estimates on the demand available before this information is received.

- Case 4 is equivalent to the uncertainty set from Ben-Tal et al. (2004) with exact revealed information and the information sets being $\{1, \ldots, t-2\}$.

To compare the solutions in different cases we have to take into account that there could be multiple optimal solutions. These solutions all give the same worst case costs, but could perform differently on individual demand trajectories and therefore also result in different mean costs. To overcome this problem, we used the two step approach that has been given in Iancu and Trichakis (2013) and Chapter 5. In this two step approach, one first minimizes the worst case costs as usual in robust optimization. To choose one solution among the set of robustly optimal solutions that performs good on average, a second step is introduced. In this second step, we add a constraint that the worst case costs do not exceed the optimal worst case costs and we replace the objective by the costs attained for the nominal demand. If in the second step the costs are minimized for the nominal demand, then one obtains the costs that are best for the mean.

The mean costs in Table 6.2 show a strange pattern among the different cases at first sight. For instance, Case 5 produces higher mean costs than Case 6, but the estimation error is much less. This phenomenon can be explained in the following way. In the two step approach, we first search for a solution with minimal worst case costs F^* and then we search among all solutions with worst case costs F^* for the solution that minimizes the nominal demand trajectory. Hence, the information in Case 2 is used to decrease the worst case costs, possibly at the costs of the average behavior.

6.5.2 Comparison with affinely adjustable robust model based on exact data

For each case we compare the WC costs and feasibility of the ARCID to the costs and feasibility resulting from the AARC approach, where one is only allowed to use the estimates that are exact (estimates with an estimation error of 0%). Hence, for the AARC solutions we only included the exact estimates, those corresponding with $\rho_{r,s} = 0$, in the decision rule. The results are given in Table 6.3.

Case 4 only deals with exact estimates. The ARCID and the AARC are equivalent in those cases because there is no estimation uncertainty. There are other situations,

Table 6.3 – WC costs of the AARC model and the ARCID model for each case.

	Worst case costs	
Cases	AARC	ARCID
1	44,273	44,268
2	44,273	44,273
3	44,582	44,582
4	44,582	44,582
5	Infeasible	44,883
6	Infeasible	45,326

namely in Case 5 and 6, where the ARCID use the extra inexact data to produce feasible solutions whereas the AARC is infeasible.

For the cases where both the AARC and the ARCID model are feasible, we notice that there is only a minor improvement in the worst case costs. For those cases, the question might rise whether we can neglect the estimation error and just apply the AARC model from Ben-Tal et al. (2004). In contrast to the AARC that we used to obtain the results in Table 6.3, we now take the information set for the AARC that includes all (estimated) demands that have an estimation error less than 100%. Hence, all estimation errors strictly between 0% and 100% are neglected and the corresponding demand estimates are used as if they were exact. To empirically see how many violations occur if the inexact nature is neglected in the AARC model, we also have to draw the demand estimates in each of the 100 demand trajectories. We draw the estimates on demand from a uniform distribution as well, using the same simulated actual demand trajectories across all cases. In every period t we know for the estimate $\widehat{d}_{r,t}$ on the simulated demand in period r that $\widehat{d}_{r,t} - d_r \in [-\rho_{r,t}\theta d_r^*, \rho_{r,t}\theta d_r^*]$, where the value d_r is taken from the earlier simulated demand patterns. Furthermore, $\widehat{d}_{r,t}$ resides in the box region $[(1-\theta)d_r^*, (1+\theta)d_r^*]$. The estimates are therefore uniformly drawn from the region:

$$[d_r - \rho_{r,t}\theta d_r^*, d_r + \rho_{r,t}\theta d_r^*] \cap [(1-\theta)d_r^*, (1+\theta)d_r^*].$$

For each case we check for how many demand trajectories, out of the 100 simulated realizations, the inventory level is lower than the minimum inventory level $V_{\min}$ of 500 or higher than the maximum inventory level $V_{\max}$ at some point in the planning period. The results are given in Table 6.4.

In Case 4 there are no violations, since this one is equivalent to the AARC based on exact information as we argued in Section 6.5.1. Table 6.4 also shows that constraints are violated more often when the estimation uncertainty is in the recent periods t

Table 6.4 – Percentage of simulated demand trajectories that violate the minimum required inventory level ($V_{\min}$) and maximum allowed inventory level ($V_{\max}$) when neglecting estimation errors.

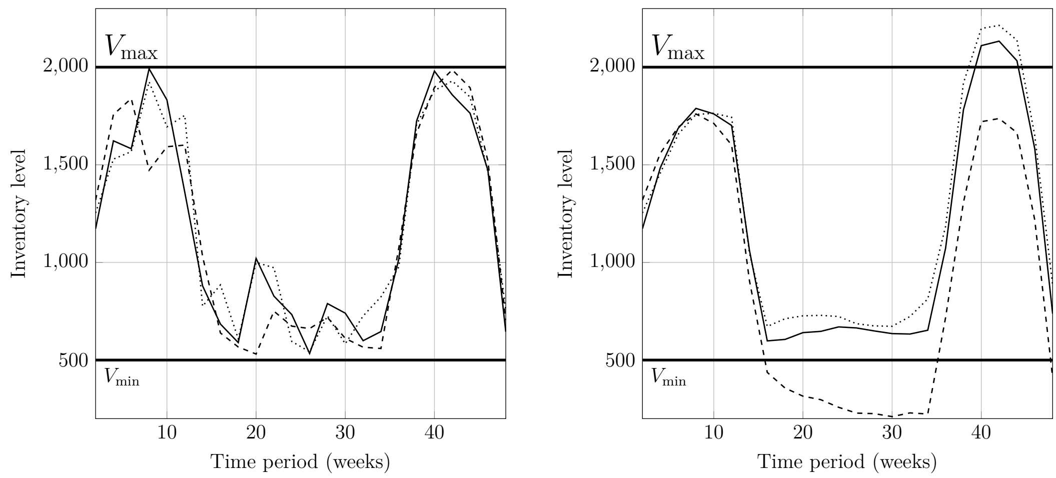
	Percentage of demand trajectories that violate the bounds	
Cases	$V_{\min}$	$V_{\max}$
1	64	55
2	80	38
3	42	38
4	0	0
5	27	15
6	26	15

and $t-1$. For example, the solution in Case 1, which has only 10% estimation uncertainty in period t, violates the minimum required inventory level 64 out of 100 times and for 55 simulated demand trajectories the stock level exceeded maximum allowed inventory level. The inventory levels for three arbitrary trajectories of Case 1 are depicted in Figure 6.4 for both the ARCID and the AARC that neglects the estimation errors.

6.6 Conclusions

In this study we consider uncertain multistage inventory systems where the observed data on demand obtained in each period is inexact. We extend the adjustable robust counterpart (ARC) method for production-inventory problems to the (ARCID) model in which the decision rules are based on *inexact* revealed data. Our numerical results demonstrate that ARCID outperforms ARC, which can only rely on exact revealed demand data. Two cases that are infeasible for the ARC solution, are feasible for the ARCID model. It is evident that neglecting the inexact nature of the revealed data may have severe consequences. For example, the inventory level dropped below the allowed minimum in up to 80% of the simulated demand trajectories.

The use of the ARCID method is thus well justified, since the resulting optimization problem that need to be solved maintains a comparable tractability status to that of the ARC method. Furthermore, there exist several software packages, such as YALMIP (Löfberg 2004), ROME (Goh and Sim 2011) and AIMMS (Roelofs and Bisschop 2012), that can do reformulation of adjustable robust optimization problems which can be readily extended to the ARCID model. Finally, we emphasize that the



(a) ARCID solution. **(b)** AARC solution.

Figure 6.4 – Inventory level of Case 1 for three simulated demand trajectories when estimation errors are taken into account (ARCID) and when estimation errors are neglected (AARC).

ARCID model set up in this chapter can also be applied to other ARC models where revealed data in each stage is inexact in various areas of operations management, such as facility location planning, flexible commitment models, capacity expansion planning, portfolio optimization and management of power systems.

6.A Proof of Theorem 6.2

Proof. We can replace the semi-infinite constraint by constraints involving maximization over the uncertainty and obtain the following constraint:

$$\mathbf{a}_i{}^\top \mathbf{x} + \sum_{t=1}^{T} \boldsymbol{b}_{it}^\top \mathbf{u}_t + \max_{(\zeta, \widehat{\zeta}^1, \widehat{\zeta}^2, \ldots, \widehat{\zeta}^T) \in \mathcal{U}_T} \left\{ \begin{pmatrix} \mathbf{A}_i^\top \mathbf{x} \\ (\mathbf{V}^1 \boldsymbol{R}^1)^\top \boldsymbol{b}_{i1} \\ \vdots \\ (\mathbf{V}^T \boldsymbol{R}^T)^\top \boldsymbol{b}_{iT} \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\zeta} \\ \widehat{\boldsymbol{\zeta}}^1 \\ \vdots \\ \widehat{\boldsymbol{\zeta}}^T \end{pmatrix} \right\} \le d_T,$$

or, by using the definition of support functions,

$$\mathbf{a}_i{}^\top \mathbf{x} + \sum_{t=1}^{T} \boldsymbol{b}_{it}^\top \mathbf{u}_t + \delta^* \left(\begin{pmatrix} \mathbf{A}_i^\top \mathbf{x} \\ (\mathbf{V}^1 \boldsymbol{R}^1)^\top \boldsymbol{b}_{i1} \\ \vdots \\ (\mathbf{V}^T \boldsymbol{R}^T)^\top \boldsymbol{b}_{iT} \end{pmatrix} \middle| \, \mathcal{U}_T \right) \le d_T, \tag{6.9}$$

Hence, all we need to do is to find an expression for the support function, similar as we did in the proof of Theorem 6.1. To do so, note that for the indicator function we have now

$$\delta\left(\left.\begin{pmatrix}\boldsymbol{\zeta}\\ \widehat{\boldsymbol{\zeta}}^1\\ \vdots\\ \widehat{\boldsymbol{\zeta}}^T\end{pmatrix}\,\right|\,\mathcal{U}\right) = \delta(\boldsymbol{\zeta} \mid \mathcal{Z}) + \sum_{t=1}^{T}\delta(\widehat{\boldsymbol{\zeta}}^t \mid \mathcal{Z}_t) + \sum_{t=1}^{T}\delta\left((\widehat{\boldsymbol{\zeta}}^t - \boldsymbol{\zeta}) \mid \widehat{\mathcal{Z}}_t\right).$$

If we define the function $h_t(\boldsymbol{\zeta}, \widehat{\boldsymbol{\zeta}}^t) = \delta\left((\widehat{\boldsymbol{\zeta}}^t - \boldsymbol{\zeta}) \mid \widehat{\mathcal{Z}}_t\right)$, then by using the definition of conjugate functions we obtain

$$h_t^*(\boldsymbol{w}_{it}, \widetilde{\boldsymbol{w}}_{it}) = \begin{cases}\delta^*\left(\boldsymbol{w}_{it} \mid \widehat{\mathcal{Z}}_t\right) & \text{if } \widetilde{\boldsymbol{w}}_{it} + \boldsymbol{w}_{it} = 0\\ \infty & \text{otherwise.}\end{cases}$$

Using this conjugate function, and the fact that $\mathcal{U}$ has nonempty relative interior, we can now find the expression for the support function in (6.3) using the sum relation for conjugate functions (see again Rockafellar (1970, Chapter 16)):

$$\delta^*\left(\left.\begin{pmatrix}\mathbf{A}_i^\top \mathbf{x}\\ (\mathbf{V}^1\boldsymbol{R}^1)^\top \boldsymbol{b}_{i1}\\ \vdots\\ (\mathbf{V}^T\boldsymbol{R}^T)^\top \boldsymbol{b}_{iT}\end{pmatrix}\,\right|\,\mathcal{U}_T\right) =$$

$$= \min_{\boldsymbol{w}_i, \widetilde{\boldsymbol{w}}_i, \boldsymbol{z}_i, \widetilde{\boldsymbol{z}}_i}\Bigg\{\delta^*(\boldsymbol{z}_i \mid \mathcal{Z}) + \sum_{t=1}^{T}\delta^*(\widetilde{\boldsymbol{z}}_{it} \mid \mathcal{Z}) + \sum_{t=1}^{T}\delta^*(\boldsymbol{w}_{it} \mid \widehat{\mathcal{Z}}) \mid \boldsymbol{z}_i + \sum_{t=1}^{T}\boldsymbol{w}_{it} = \mathbf{A}_i^\top \mathbf{x},$$

$$\widetilde{\boldsymbol{w}}_{it} + \widetilde{\boldsymbol{z}}_{it} = (\mathbf{V}^t\boldsymbol{R}^t)^\top \boldsymbol{b}_{it},\ \boldsymbol{w}_{it} + \widetilde{\boldsymbol{w}}_{it} = 0 \quad \forall t = 1, \ldots, T\Bigg\}.$$

Substituting this result into (6.9) yields that (6.4) is feasible if and only if there exist $\boldsymbol{z}_i$, $\widetilde{\boldsymbol{z}}_{i1}, \ldots, \widetilde{\boldsymbol{z}}_{iT}$, $\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iT}$, $\widetilde{\boldsymbol{w}}_{i1}, \ldots, \widetilde{\boldsymbol{w}}_{iT} \in \mathbb{R}^L$ that satisfy

$$\begin{cases}\mathbf{a}_i{}^\top \mathbf{x} + \sum_{t=1}^{T}(\boldsymbol{b}_i^t)^\top \mathbf{u}_t + \delta^*(\boldsymbol{z}_i \mid \mathcal{Z}) + \sum_{t=1}^{T}\delta^*(\widetilde{\boldsymbol{z}}_{it} \mid \mathcal{Z}) + \sum_{t=1}^{T}\delta^*(\boldsymbol{w}_{it} \mid \widehat{\mathcal{Z}}) \le d_i\\ \boldsymbol{z}_i + \sum_{t=1}^{T}\boldsymbol{w}_{it} = \mathbf{A}_i^\top \mathbf{x}\\ \widetilde{\boldsymbol{w}}_{it} + \widetilde{\boldsymbol{z}}_{it} = (\mathbf{V}^t\boldsymbol{R}^t)^\top \boldsymbol{b}_{it} \qquad \forall t = 1, \ldots, T\\ \boldsymbol{w}_{it} + \widetilde{\boldsymbol{w}}_{it} = 0 \qquad \forall t = 1, \ldots, T.\end{cases}$$

The result then follows by elimination of the variables $\widetilde{\boldsymbol{w}}_{it}$, $\widetilde{\boldsymbol{z}}_{it}$ for all $t = 1, \ldots, T$ and $\boldsymbol{z}_i$. ■

6.B The tractable robust counterpart based on inexact data

Here we present the final tractable robust counterpart for the model (P:ARCID). Note that all but the last two sets of constraints on $V_{\min}$ and $V_{\max}$ are the same as in Ben-Tal et al. (2004), since those are the only constraints involving both the true demand parameters and their inexact estimates.

$$
\begin{aligned}
&\min_{\pi,F,\alpha,\beta,\gamma,\delta,\zeta,\xi,\eta,\mu,\nu} \quad F\\
&s.t.\\
&\sum_{t=1}^{T}\sum_{i=1}^{I} c_i(t)\pi_{i,t}^0 + \sum_{r=1}^{T}\alpha_r d_r^* + \theta\sum_{r=1}^{T}\beta_r d_r^* \le F\\
&\sum_{i=1}^{I}\sum_{t:r\in\hat{I}_t} c_i(t)\pi_{i,t}^r = \alpha_r,\ -\beta_r \le \alpha_r \le \beta_r, \quad 1\le r\le T\\
&-\gamma_{i,t}^r \le \pi_{i,t}^r \le \gamma_{i,t}^r, \qquad 1\le i\le I, \qquad 1\le r,t\le T;\\
&\pi_{i,t}^0 + \sum_{r\in\hat{I}_t}\pi_{i,t}^r d_r^* - \theta\sum_{r\in\hat{I}_t}\gamma_{i,t}^r d_r^* \ge 0, \qquad 1\le i\le I,\ 1\le t\le T,\\
&\pi_{i,t}^0 + \sum_{r\in\hat{I}_t}\pi_{i,t}^r d_r^* + \theta\sum_{r\in\hat{I}_t}\gamma_{i,t}^r d_r^* \le P_i(t), \qquad 1\le i\le I,\ 1\le t\le T,\\
&\sum_{t:r\in\hat{I}_t}\pi_{i,t}^r = \delta_i^r, -\zeta_i^r \le \delta_i^r \le -\zeta_i^r, \quad 1\le r\le T\\
&\sum_{t=1}^{T}\pi_{i,t}^0 + \sum_{r=1}^{T}\delta_i^r d_r^* + \theta\sum_{r=1}^{T}\zeta_i^r d_r^* \le Q_i, \qquad 1\le i\le I\\
&\tau_t^r = -1 - \sum_{s\le t,r\in\hat{I}_t}\lambda_{s,t}^r,\ \xi_t^r = \sum_{s\le t,r\in\hat{I}_t}\sum_{i=1}^{I}\pi_{i,s}^r - 1, \qquad 1\le r\le t\le T\\
&-\mu_{s,t}^r \le \sum_{i=1}^{I}\pi_{i,s}^r + \lambda_{s,t}^r \le \mu_{s,t}^r,\ -\omega_{s,t}^r \le \lambda_{s,t}^r \le \omega_{s,t}^r, \quad s: r\in I_s, \qquad 1\le r\le t\le T\\
&-\nu_t^r \le \tau_t^r \le \nu_t^r,\ \eta_t^r = \nu_t^r + \sum_{s=1}^{t}\mu_{s,t}^r, \qquad 1\le r\le t\le T\\
&\sum_{s=1}^{t}\sum_{i=1}^{I}\pi_{i,s}^0 + \sum_{r=1}^{t}\xi_t^r d_r^* + \theta\sum_{r=1}^{t}\eta_t^r d_r^* + \sum_{r=1}^{t}\sum_{s\le t,r\in\hat{I}_t}\rho_{r,s}\theta\omega_{s,t}^r d_r^* \le V_{\max} - v(1), \quad 1\le t\le T\\
&-\sum_{s=1}^{t}\sum_{i=1}^{I}\pi_{i,s}^0 - \sum_{r=1}^{t}\xi_t^r d_r^* + \theta\sum_{r=1}^{t}\eta_t^r d_r^* + \sum_{r=1}^{t}\sum_{s\le t,r\in\hat{I}_t}\rho_{r,s}\theta\omega_{s,t}^r d_r^* \le v(1) - V_{\min}, \quad 1\le t\le T.
\end{aligned}
\tag{ARCID-BT}
$$

Bibliography

Adida, E. and G. Perakis (2006). A robust optimization approach to dynamic pricing and inventory control with no backorders. *Mathematical Programming* 107 (1), pp. 97–129.

Akyildiz, I., W. Su, Y. Sankarasubramaniam, and E. Cayirci (2002). Wireless sensor networks: a survey. *Computer Networks* 38 (4), pp. 393 –422.

Ardestani-Jaafari, A. and E. Delage (2016a). Linearized robust counterparts of two-stage robust optimization problem with applications in operations management. *Working paper*. Available on: www.optimization-online.org/DB_HTML/2016/03/5388.html.

Ardestani-Jaafari, A. and E. Delage (2016b). Robust optimization of sums of piece-wise linear functions with application to inventory problems. *Operations Research* 64 (2), pp. 474–494.

Ardestani-Jaafari, A. and E. Delage (2017). The value of flexibility in robust location–transportation problems. *Transportation Science*. Published online in Articles in Advance, pp. 1–21.

Atamtürk, A. and M. Zhang (2007). Two-stage robust network flow and design under demand uncertainty. *Operations Research* 55 (4), pp. 662–673.

Baron, O., J. Milner, and H. Naseraldin (2011). Facility location: a robust optimization approach. *Production and Operations Management* 20 (5), pp. 772–785.

Bemporad, A., F. Borrelli, and M. Morari (2003). Min-max control of constrained uncertain discrete-time linear systems. *IEEE Transactions on Automatic Control* 48 (9), pp. 1600–1606.

Ben-Ameur, W. (2007). Between fully dynamic routing and robust stable routing. *2007 6th International Workshop on Design and Reliable Communication Networks*, pp. 1–6.

Ben-Tal, A. and D. den Hertog (2014). Hidden conic quadratic representation of some nonconvex quadratic optimization problems. *Mathematical Programming* 143 (1-2), pp. 1–29.

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

Ben-Tal, A. and A. Nemirovski (1999). Robust solutions of uncertain linear programs. *Operations Research Letters* 25 (1), pp. 1–13.

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

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

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

Ben-Tal, A., S. Boyd, and A. Nemirovski (2006). Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming* 107 (1-2), pp. 63–89.

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

Ben-Tal, A., D. den Hertog, and M. Laurent (2011a). Hidden convexity in partially separable optimization. *CentER Working Paper Series No 2011-070.*

Ben-Tal, A., B. Do Chung, S. R. Mandala, and T. Yao (2011b). Robust optimization for emergency logistics planning: risk mitigation in humanitarian relief supply chains. *Transportation Research Part B: Methodological* 45 (8), pp. 1177–1189.

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

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

Ben-Tal, A., O. El Housni, and V. Goyal (2016). A tractable approach for designing piecewise affine policies in dynamic robust optimization. *Working paper*. Available on: http://www.optimization-online.org/DB_HTML/2016/07/5557.html.

Ben-Tal, A., R. Brekelmans, D. den Hertog, and J.-Ph. Vial (2017). Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing* 29 (2), pp. 350–366.

Bertsimas, D. and H. Bidkhori (2015). On the performance of affine policies for two-stage adaptive optimization: a geometric perspective. *Mathematical Programming* 153 (2), pp. 577–594.

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

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

Bertsimas, D. and M. S. Copenhaver (2017). Characterization of the equivalence of robustification and regularization in linear and matrix regression. *To appear in European Journal of Operational Research.*

Bertsimas, D. and I. Dunning (2016). Multistage robust mixed-integer optimization with adaptive partitions. *Operations Research* 64 (4), pp. 980–998.

Bertsimas, D. and V. Goyal (2010). On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research* 35 (2), pp. 284–305.

Bertsimas, D. and V. Goyal (2012). On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming* 134 (2), pp. 491–531.

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

Bertsimas, D., D. A. Iancu, and P. A. Parrilo (2010). Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research* 35 (2), pp. 363–394.

Bertsimas, D., D. A. Iancu, and P. A. Parrilo (2011a). A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control* 56 (12), pp. 2809–2824.

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

Bertsimas, D., E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems* 28 (1), pp. 52–63.

Bertsimas, D., I. Dunning, and M. Lubin (2016). Reformulation versus cutting-planes for robust optimization. *Computational Management Science* 13 (2), pp. 195–217.

Bertsimas, D., V. Gupta, and N. Kallus (2017a). Data-driven robust optimization. *To appear in Mathematical Programming.*

Bertsimas, D., M. Sim, and M. Zhang (2017b). Distributionally adaptive optimization. *To appear in Management Science.*

Birge, J. R. and F. Louveaux (2011). *Introduction to Stochastic Programming.* Springer Science & Business Media.

Boni, O. and A. Ben-Tal (2008). Adjustable robust counterpart of conic quadratic problems. *Mathematical Methods of Operations Research* 68 (2), pp. 211–233.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization.* Cambridge University Press.

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

Calafiore, G. C. (2008). Multi-period portfolio optimization with linear control policies. *Automatica* 44 (10), pp. 2463 –2473.

Calafiore, G. C. (2009). An affine control method for optimal dynamic asset allocation with transaction costs. *SIAM Journal on Control and Optimization* 48 (4), pp. 2254–2274.

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

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

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

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

Dacorogna, B. and P. Maréchal (2007). The role of perspective functions in convexity, polyconvexity, rank-one convexity and separate convexity. *Journal of Convex Analysis* 15 (2), pp. 271–284.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science* 1 (3-4), pp. 197–206.

Dantzig, G. B. (1991). Linear programming. *History of Mathematical Programming– A Collection of Personal Reminiscences*. Ed. by A. K. Lenstra, A. H. G. Rinnooy Kan, and A. Schrijver. Amsterdam: Elsevier. Chap. 3, pp. 19–31.

DeHoratius, N. and A. Raman (2008). Inventory record inaccuracy: an empirical analysis. *Management Science* 54 (4), pp. 627–641.

Delage, E. and D. A. Iancu (2015). Robust multistage decision making. *The Operations Research Revolution*. Chap. 2, pp. 20–46.

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

Dunning, I., J. Huchette, and M. Lubin (2017). JuMP: a modeling language for mathematical optimization. *SIAM Review* 59 (2), pp. 295–320.

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

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

Fischetti, M. and M. Monaci (2009). Light robustness. *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*. Ed. by R. K. Ahuja, R. H. Möhring, and C. D. Zaroliagis. Berlin, Heidelberg: Springer, pp. 61–84.

Fischetti, M. and M. Monaci (2012). Cutting plane versus compact formulations for uncertain (integer) linear programs. *Mathematical Programming Computation* 4 (3), pp. 239–273.

Gabrel, V., M. Lacroix, C. Murat, and N. Remli (2014a). Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics* 164 (1), pp. 100–111.

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

Georghiou, A., W. Wiesemann, and D. Kuhn (2015). Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming* 152 (1-2), pp. 301–338.

Goh, J. and M. Sim (2010). Distributionally robust optimization and its tractable approximations. *Operations Research* 58 (4), pp. 902–917.

Goh, J. and M. Sim (2011). Robust optimization made easy with ROME. *Operations Research* 59 (4), pp. 973–985.

Gorissen, B. L. and D. den Hertog (2013). Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research* 227 (1), pp. 30–43.

Gorissen, B. L., İ. Yanıkoğlu, and D. den Hertog (2015). A practical guide to robust optimization. *Omega* 53, pp. 124–137.

Gounaris, C. E., W. Wiesemann, and C. A. Floudas (2013). The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research* 61 (3), pp. 677–693.

Grant, M., S. Boyd, and Y. Ye (2006). Disciplined convex programming. *Global Optimization: From Theory to Implementation*. Ed. by L. Liberti and N. Maculan. Boston, MA: Springer US, pp. 155–210.

Guigues, V. and C. Sagastizábal (2012). The value of rolling-horizon policies for risk-averse hydro-thermal planning. *European Journal of Operational Research* 217 (1), pp. 129–140.

Gurobi Optimization, Inc. (2015). *Gurobi Optimizer Reference Manual*. http://www.gurobi.com.

Guslitzer, E. (2002). Uncertainty-Immunized Solutions in Linear Programming. MSc thesis. Technion-Israel Institute of Technology.

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

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

Hanasusanto, G. A., D. Kuhn, and W. Wiesemann (2015b). K-adaptability in two-stage robust binary programming. *Operations Research* 63 (4), pp. 877–891.

Haug, A., F. Zachariassen, and D. van Liempd (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management* 4 (2), pp. 168–193.

Iancu, D. A. and N. Trichakis (2013). Pareto efficiency in robust optimization. *Management Science* 60 (1), pp. 130–147.

Iancu, D. A., M. Sharma, and M. Sviridenko (2013). Supermodularity and affine policies in dynamic robust optimization. *Operations Research* 61 (4), pp. 941–956.

Kall, P. and S. W. Wallace (1994). *Stochastic Programming*. Chichester: John Wiley & Sons UK.

Kök, A. G. and K. H. Shang (2007). Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing & Service Operations Management* 9 (2), pp. 185–205.

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

Lobo, M. S., L. Vandenberghe, S. Boyd, and H. Lebret (1998). Applications of second-order cone programming. *Linear Algebra and its Applications* 284 (1-3), pp. 193–228.

Löfberg, J. (2004). YALMIP–a toolbox for modeling and optimization in MATLAB. *Proceedings of the CACSD Conference*. Taipei, Taiwan.

Lubin, M. and I. Dunning (2015). Computing in operations research using Julia. *INFORMS Journal on Computing* 27 (2), pp. 238–248.

Minoux, M. (2011). On 2-stage robust LP with RHS uncertainty: complexity results and applications. *Journal of Global Optimization* 49 (3), pp. 521–537.

MOSEK ApS. (2017). *The MOSEK optimization toolbox for MATLAB manual*. http://docs.mosek.com/8.0/toolbox.pdf.

Mutapcic, A. and S. Boyd (2009). Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software* 24 (3), pp. 381–406.

Ng, T. S. and C. Sy (2014). An affine adjustable robust model for generation and transmission network planning. *International Journal of Electrical Power & Energy Systems* 60, pp. 141–152.

Ordóñez, F. and J. Zhao (2007). Robust capacity expansion of network flows. *Networks* 50 (2), pp. 136–145.

Pınar, M. Ç. and R. H. Tütüncü (2005). Robust profit opportunities in risky financial portfolios. *Operations Research Letters* 33 (4), pp. 331–340.

Poss, M. (2014). Robust combinatorial optimization with variable cost uncertainty. *European Journal of Operational Research* 237 (3), pp. 836–845.

Postek, K. and D. den Hertog (2016). Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing* 28 (3), pp. 553–574.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM* 41 (2), pp. 79–82.

Rocha, P. and D. Kuhn (2012). Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *European Journal of Operational Research* 216 (2), pp. 397–408.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

Roelofs, M. and J. Bisschop (2012). *AIMMS 4.2: the language reference*. Paragon Decision Technology BV.

See, C.-T. and M. Sim (2010). Robust approximation to multiperiod inventory management. *Operations Research* 58 (3), pp. 583–594.

Shapiro, A. and A. Nemirovski (2005). On complexity of stochastic programming problems. *Continuous Optimization: Current Trends and Modern Applications*. Ed. by V. Jeyakumar and A. Rubinov. Boston, MA: Springer US, pp. 111–146.

Soffer, P. (2010). Mirror, mirror on the wall, can I count on you at all? Exploring data inaccuracy in business processes. *Enterprise, Business-Process and Information Systems Modeling*. Springer, pp. 14–25.

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

Takeda, A., S. Taguchi, and R. H. Tütüncü (2008). Adjustable robust optimization models for a nonlinear two-period system. *Journal of Optimization Theory and Applications* 136 (2), pp. 275–295.

Thiele, A., T. Terry, and M. Epelman (2009). Robust linear optimization with recourse. *Working paper*. Available on: http://www.optimization-online.org/DB_FILE/2009/03/2263.pdf, pp. 4–37.

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

Wang, Q., J.-P. Watson, and Y. Guan (2013). Two-stage robust optimization for Nk contingency-constrained unit commitment. *IEEE Transactions on Power Systems* 28 (3), pp. 2366–2375.

Wiesemann, W., D. Kuhn, and B. Rustem (2012). Robust resource allocations in temporal networks. *Mathematical Programming* 135 (1), pp. 437–471.

Wisenhausen, H. (1966). Minimax Control of Uncertain Systems. PhD thesis. Massachusetts Institute of Technology.

Xu, H., C. Caramanis, and S. Mannor (2009). Robust regression and lasso. *Advances in Neural Information Processing Systems* 56 (7), pp. 1801–1808.

Yin, Y., S. M. Madanat, and X.-Y. Lu (2009). Robust improvement schemes for road networks under demand uncertainty. *European Journal of Operational Research* 198 (2), pp. 470–479.

Zeng, B. and L. Zhao (2013). Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters* 41 (5), pp. 457–461.

Zhao, L. and B. Zeng (2012). Robust unit commitment problem with demand response and wind energy. *2012 IEEE Power and Energy Society General Meeting*, pp. 1–8.

Zhen, J. and D. den Hertog (2017). Computing the maximum volume inscribed ellipsoid of a polytopic projection. *To appear in INFORMS Journal on Computing.*

Zhen, J., D. den Hertog, and M. Sim (2016). Adjustable robust optimization via Fourier-Motzkin elimination. *Working paper*. Available on: www.optimization-online.org/DB_FILE/2016/07/5564.pdf.