Dual approach for two-stage robust nonlinear optimization

Frans J.C.T. de Ruiter
CQM, Eindhoven, The Netherlands
Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands

Jianzhe Zhen*
Risk Analytics and Optimization Chair, École Polytechnique Fédérale de Lausanne, Switzerland
Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands

Dick den Hertog
Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands

Adjustable robust optimization models in which the adjustable variables appear in a convex way are difficult to solve. For example, if we substitute linear decision rules for the adjustable variables, then the model becomes convex in the uncertain parameters, whereas for computational tractability we need concavity in the uncertain parameters. These linear decision rules, as well as other techniques like Fourier-Motzkin elimination of the adjustable variables all require that these variables appear in a linear way.

In this paper we reformulate the original adjustable robust nonlinear problem with a polyhedral uncertainty set into an equivalent adjustable robust linear problem, for which approaches as Fourier-Motzkin elimination and (non)linear decision rules can be used. The reformulation is obtained by first dualizing over the adjustable variables and then over the uncertain parameters. The polyhedral structure of the uncertainty set then appears in the linear constraints of the dualized problem, and the nonlinear functions of the adjustable variables in the original problem appear in the uncertainty set of the dualized problem. This paper also describes how to effectively obtain lower bounds on the optimal objective value by linking the realizations in the uncertainty set of dualized problem to realizations in the original uncertainty set. Two numerical examples, one on a distribution problem on a network with commitments, and one on finding the equilibrium of a system with piecewise-linear springs, show the effectiveness and applicability of the new approach.

Key words: Adjustable robust optimization, nonlinear inequalities, duality, linear decision rules.

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, 1999, 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

* Corresponding author: jianzhe.zhen@epfl.ch

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 (also known as adjustable variables) 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 solved using adjustable robust linear optimization such as inventory models (Ben-Tal et al., 2004, 2005), facility location planning (Ardestani-Jaafari and Delage, 2017; Atamtürk and Zhang, 2007; Gabrel et al., 2014), energy production scheduling (Bertsimas et al., 2013; Ng and Sy, 2014), project management (Wiesemann et al., 2012), portfolio optimization (Calafiore, 2008, 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, for adjustable robust linear optimization models with fixed recourse, good solutions can be found using linear decision rules. Rather than allowing the adjustable variable 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 coefficients of the linear 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). There are also several other papers that establish optimality or give theoretical a-priori bounds on the objective value (Bertsimas and Bidkhori, 2015; Bertsimas and Goyal, 2012). Another recently developed method, that is also used in this chapter, is Fourier-Motzkin elimination for adjustable robust optimization. This method can solve small adjustable robust linear optimization models to optimality by eliminating the adjustable variables. For larger problems we can eliminate some of the adjustable variables 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. 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 dealing with 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 adjustable variables. To obtain a static robust model one could try to substitute a linear decision rule for the adjustable variables. 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 in a folding horizon way. 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 on adjustable robust nonlinear 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 this can be generalized to problems with more constraints, other uncertainty sets or other model formulations. Takeda et al. (2008) consider an adjustable robust nonlinear model with polyhedral uncertainty set, similar to the models considered in this paper. 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 and finally end up with a semidefinite optimization model.

In this paper we propose a computationally tractable approach for adjustable robust optimization models that are convex in the adjustable variables and that have polyhedral uncertainty sets. As a special case, our method is equivalent to Bertsimas and de Ruiter (2016) if all functions involved are linear. It was shown in that paper 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 paper we consider a more general setting, which allows us to consider models that are nonlinear in the adjustable variables. Apart from providing the first tractable way of solving the more general adjustable robust nonlinear models, we show how scenarios in the uncertainty set of the primal and dualized version are tied to each other. To summarize, our contributions in this chapter are:

1. We develop an approach for two-stage robust nonlinear problems that are fixed recourse and that have a polyhedral uncertainty set. In this approach we consecutively dualize over adjustable

variables and uncertain parameters. The resulting model that again can be interpreted as a two-stage robust problem but now with adjustable variables that appear linearly, 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. 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.

3. We show that we can use our method to efficiently solve practical two-stage robust nonlinear optimization models. This is done via two numerical experiments: distribution on a network with commitments and finding the equilibrium of a system with piecewise-linear springs. We use Fourier-Motzkin elimination in combination with 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 §2 we present our framework and derive our dualized formulation. We then extend the new approach to a more general class of models that can be reformulated to fit into our framework. In §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 §4 and §5.

Notation. The indicator function of a set $\mathcal{S} \subseteq \mathbb{R}^{n_\nu}$ is denoted by $\delta(\nu|\mathcal{S})$ and defined by:

$$\delta(\nu|\mathcal{S}) = \begin{cases} 0 & \nu \in \mathcal{S} \\ \infty & \nu \notin \mathcal{S}. \end{cases}$$

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

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

where $\mathrm{dom}(g)$ is the domain of the function g. The conjugate function of indicator function $\delta(\nu|\mathcal{S})$ is called the support function of $\mathcal{S}$ and given by:

$$\delta^*(z|\mathcal{S}) = \sup_{\nu} \left\{ \nu^\top z - \delta(\nu|\mathcal{S}) \right\} = \sup_{\nu \in \mathcal{S}} \left\{ \nu^\top z \right\}.$$

The perspective $h : \mathbb{R}^{n_\nu+1} \to \mathbb{R}$ of a closed convex function $f : \mathbb{R}^{n_\nu} \to \mathbb{R}$ is defined for all $\nu \in \mathbb{R}^{n_\nu}$ and $t \in \mathbb{R}_+$ as $h(\nu,t) = tf(\nu/t)$ if $t > 0$, and $h(\nu,0) = 0$ if $\nu = 0$, ∞ otherwise.

2. Linear dual formulation

2.1. Framework

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

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

where $\mathcal{X}\subseteq\mathbb{R}^{n_x}$, the functions $f_i:\mathbb{R}^{n_x}\to\mathbb{R}$, $g_i:\mathbb{R}^{n_y}\to\mathbb{R}$ are proper closed convex for all $i=0,\ldots,m_1$, $F_{i,\cdot}(x)=(F_{i,1}(x),\ldots,F_{i,n_\zeta}(x))$ and $F_{i,l}:\mathbb{R}^{n_x}\to\mathbb{R}$ are real valued functions for all $i=1,\ldots,m_1$, and $l=1,\ldots,n_\zeta$. 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:

$$A(\zeta)=A^0+\sum_{l=1}^{n_\zeta}A^l\zeta_l,\qquad b(\zeta)=b^0+\sum_{l=1}^{n_\zeta}b^l\zeta_l,\tag{2}$$

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

$$\mathcal{U}=\{\zeta\geq 0:D\zeta\leq d\}.\tag{3}$$

The nonnegativity constraints for ζ in (3) are just for ease of analysis, but our approach also works without these constraints. Model (1) is generally intractable, and we cannot approximate model (1) with linear decision rules as is. If we substitute linear decision rules $y=Q\zeta+q$, then the objective and constraints have terms $g_i(Q\zeta+q)$ for all $i=0,\ldots,m_1$, 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. In the next subsection, we use a dual approach to derive an equivalent reformulation of (1). The reformulation is a two-stage robust linear optimization model, for which approaches as Fourier-Motzkin elimination and (non)linear decision rules can be used.

2.2. Consecutive dualization

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 $g_0,\ldots,g_{m_1}$, such that*

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

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

$$\zeta^\top F_{i,\cdot}(x)+f_i(x)+g_i(y)<0.$$

$\square$

This assumption implies that each here-and-now decision is strictly feasible. This assumption is required to guarantee strong duality by Slaters' condition in Theorem 1. 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 $g_0(y)$. Also, 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 (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 (1) is more flexible, e.g., functions $g(x,y,\zeta)$ also fit into the framework for special structures by introducing additional adjustable variables and constraints as shown in §2.4.

In the following theorem, we reformulate (1) via an approach that we call *consecutive dualization.* We first dualize (1) over the adjustable variable y to obtain a *inf-sup-sup* model and then consecutively dualize over the uncertain parameter ζ. This approach and the resulting dualized formulation are formally described in the following theorem and proof.

Theorem 1. *Let* $\mathcal{U}$ *be a polyhedral set as in* (3) *and assume that Assumption 1 holds. The here-and-now decision* x *is feasible for* (1) *if and only if* x *is feasible for the following dualized model:*

$$\inf_{x\in\mathcal{X}}\sup_{(u,v,w,z)\in\mathcal{V}}\inf_{\lambda\geq 0}\left\{\sum_{i=0}^{m_1}v_if_i(x)+d^\top\lambda+w^\top\left(A^0x-b^0\right)-\sum_{i=0}^{m_1}z_i\;\middle|\;\sum_{j=1}^{p}D_{j,l}\lambda_j\geq w^\top(A^lx-b^l)+\sum_{i=1}^{m_1}v_iF_{i,l}(x),\quad l=1,\ldots,n_\zeta\right\},\tag{4}$$

where $u=(u_0,\ldots,u_{m_1})\in\mathbb{R}^{(m_1+1)n_y}$, $u_i\in\mathbb{R}^{n_y}$ *for* $i=0,...,m_1$, *and*

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

Proof. We consider the inner infimum of (1) over y for a given $x\in\mathcal{X}$ and $\zeta\in\mathcal{U}$. Since Assumption 1 holds we can apply the Lagrangian principle to (1), and obtain the following equivalent reformulation:

$$\inf_{x\in\mathcal{X}}\sup_{\substack{\zeta\in\mathcal{U}\\ v\geq 0,w}}\inf_{y}\ f_0(x)+g_0(y)+\sum_{i=1}^{m_1}v_i\left(\zeta^\top F_{i,\cdot}(x)+f_i(x)+g_i(y)\right)+w^\top\left(A(\zeta)x+By-b(\zeta)\right).\tag{5}$$

We then use the definition of the conjugate functions and calculus rules for conjugate functions (specifically Rule 5 in Table 2 of Roos et al. (2016) for the conjugate of the sum of convex functions) to obtain the following inf-sup-sup reformulation:

$$\inf_{x\in\mathcal{X}}\sup_{\zeta\in\mathcal{U}}\sup_{(u,v,w)\in\mathcal{W}} f_0(x)+\sum_{i=1}^{m_1} v_i\left(\zeta^\top F_{i,\cdot}(x)+f_i(x)\right)+w^\top\left(A(\zeta)x-b(\zeta)\right)-\sum_{i=0}^{m_1} v_i\left(g_i\right)^*\left(\frac{u_i}{v_i}\right), \tag{6}$$

where $\mathcal{W}=\left\{(u,v,w):\ v\geq 0,\ v_0=1,\ \sum_{i=0}^{m_1}u_i=-B^\top w\right\}$. We can then switch the order of supremum such that the inner supremum is over $\zeta\in\mathcal{U}$. Since the inner supremum model is linear in ζ, we can apply strong duality for linear optimization to obtain the inf-sup-inf reformulation:

$$\inf_{x\in\mathcal{X}}\sup_{(u,v,w)\in\mathcal{W}}\inf_{\lambda\geq 0}\left\{\sum_{i=0}^{m_1} v_i f_i(x)+d^\top\lambda+w^\top\left(A^0x-b^0\right)-\sum_{i=0}^{m_1} v_i\left(g_i\right)^*\left(\frac{u_i}{v_i}\right)\ \middle|\ \sum_{j=1}^{p} D_{j,l}\lambda_j\geq w^\top(A^l x-b^l)+\sum_{i=1}^{m_1} v_i F_{i,l}(x),\quad l=1,\ldots,n_\zeta\right\}.$$

We then introduce epigraph variables z_i for every $v_i\left(g_i\right)^*\left(\frac{u_i}{v_i}\right)$, $i=0,...,m_1$, and finally obtain (4). □

Note that the linear structure of the uncertainty set appears in the constraints of the dual formulation (4) and the convex structure of the adjustable variables is in the new uncertainty set of (4). In the specific linear case, where $F_{i,\cdot}(x), f_i(x)$ and $g_i(y)$ are affine functions, Theorem 1 coincides with the result in (Bertsimas and de Ruiter, 2016, Theorem 1). The main benefit (and purpose) of dualization is that the resulting model is linear in the adjustable variables and uncertain parameters, and has fixed recourse, which 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. For several linear two-stage robust models, the structure of the optimal decision rules has been characterized. For instance, by using Zhen and den Hertog (2017), it can be shown that there exists a polynomial of (at most) degree $(m_1+1)(n_y+2)+m_2$ and linear in $u_i, v_i,\ i=0,\ldots,m_1$, and $w_i,\ i=1,\ldots,m_2$, that is optimal for λ in (4). One can use Zhen et al. (2017b) to establish that there exists a piecewise affine function that is optimal for λ in (4). More specifically, if $\mathcal{U}$ is simplicial, there exists a linear decision rule that is optimal for λ in (4); if $\mathcal{U}$ is a box, there exists a two-piecewise affine function that is optimal for λ in (4), and the techniques proposed in Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016) can then be applied to solve Problem (4) approximately. Unfortunately, even when the structure of optimal decision rules is known, it is often hard to find optimal solutions due to the computational intractability of such rules.

In many cases $\mathcal{V}$ is second-order cone representable, making (4) a second-order cone (SOC) problem if linear decision rules are used (see, e.g., Example 1), which can be efficiently solved with off-the-shelve solvers. For a comprehensive list of convex conjugate functions as well as techniques to derive them, we refer to the paper Roos et al. (2016). One may argue that it can be cumbersome or impossible to derive the closed form of the conjugates for some g_i. In §2.3, we show that after applying popular approaches such as Fourier-Motzkin elimination and linear decision rules, only the perspectives of g_i (not the conjugates of g_i) play a role in the tractable robust counterpart of (4). Many uncertainty sets naturally require $\zeta \geq 0$, which is the reason we impose it here. However, the nonnegativity restriction on ζ in (3) can be omitted. In that case one will end up with equality constraints in the first n_ζ constraints of (4). Each equality constraint can be eliminated by eliminating one of the adjustable variables that appear in that equality constraint. Note that we did not need to assume convexity for the functions f and F in model (1) for Theorem 1, but to end up with tractable models we usually assume these functions are (componentwise) convex.

2.3. Using linear decision rules for the dualized problem

In this section we show how to solve the dualized formulation (4) with linear decision rules. The adjustable variable λ in (4) is a function of the uncertain parameter $(u,v,w,z) \in \mathcal{V}$. As it is usually done in adjustable robust optimization, we use linear decision rules for λ. Most interestingly from a practitioner's perspective we find that the resulting optimization model does not contain the conjugates g_i^*, but only the original functions g_i. Substituting the linear decision rule:

$$\lambda = q + Qu + Rv + Sw + Tz$$

in (4), yields the following static robust optimization problem:

$$\begin{aligned}
&\inf_{\substack{x\in\mathcal{X},q\\ Q,R,S,T}} \sup_{(u,v,w,z)\in\mathcal{V}} v^\top F_0(x) + d^\top(q+Qu+Rv+Sw+Tz) + w^\top\left(A^0x - b^0\right) - e^\top z_i \\
&\text{s.t. } \forall (u,v,w,z)\in\mathcal{V}: \\
&\qquad \begin{cases} D_{\cdot,l}^\top(q+Qu+Rv+Sw+Tz) \geq w^\top(A^l x - b^l) + v^\top \begin{bmatrix} 0 \\ F_{\cdot,l}(x)\end{bmatrix} & l = 1,\ldots,n_\zeta \\ q+Qu+Rv+Sw+Tz \geq 0, \end{cases}
\end{aligned} \tag{7}$$

where $Q \in \mathbb{R}^{p\times(m_1+1)n_y}$, $R,\ T \in \mathbb{R}^{p\times(m_1+1)}$, $S \in \mathbb{R}^{p\times m_2}$ and $F_0(x) = (f_0(x),\ldots,f_{m_1}(x))^\top \in \mathbb{R}^{m_1+1}$, $F_{\cdot,l}(x) \in \mathbb{R}^{m_1}$ and $D_{\cdot,l} \in \mathbb{R}^p$ are the l-th column vectors of $F(x)$ and D, respectively. Since the objective and constraints are linear in all the uncertain parameters, we can rewrite (7) into the following equivalent reformulation by collecting all the terms of respectively the parameters u,v,w and z, and use the definition of the support function:

$$\begin{aligned}
\inf_{\substack{x\in\mathcal{X}\\ q,Q,R\\ S,T}} \quad & d^\top q + \delta^*\left(\left(\begin{smallmatrix} Q^\top d \\ R^\top d + F_0(x) \\ S^\top d + A^0 x - b^0 \\ T^\top d - e \end{smallmatrix}\right) \middle| \mathcal{V}\right) \\
\text{s.t.} \quad & D_{\cdot,l}^\top q \geq \delta^*\left(\left(\begin{smallmatrix} -Q^\top D_{\cdot,l} \\ \left[\begin{smallmatrix} 0 \\ F_{\cdot,l}(x)\end{smallmatrix}\right] - R^\top D_{\cdot,l} \\ A^l x - b^l - S^\top D_{\cdot,l} \\ -T^\top D_{\cdot,l} \end{smallmatrix}\right) \middle| \mathcal{V}\right) \qquad l = 1,\dots,n_\zeta \\
& q_j \geq \delta^*\left(\left(\begin{smallmatrix} -Q_{j,\cdot} \\ -R_{j,\cdot} \\ -S_{j,\cdot} \\ -T_{j,\cdot} \end{smallmatrix}\right) \middle| \mathcal{V}\right) \qquad j = 1,\dots,p,
\end{aligned} \tag{8}$$

where e is the all-one vector, $Q_{j,\cdot} \in \mathbb{R}^{(m_1+1)n_y}$, and $R_{j,\cdot}, S_{j,\cdot}, T_{j,\cdot} \in \mathbb{R}^{m_1+1}$ are the j-th row vector of matrices Q, R, S and T, respectively. A computationally tractable robust counterpart of (8) can be obtained by calculating the support function of the uncertainty set $\mathcal{V}$, using the approach of (Ben-Tal et al., 2015). In the next theorem we show that the support function of $\mathcal{V}$ has a simple format, that is, a minimization problem without conjugates g_i^*.

Theorem 2.

$$\delta^*\left(\left(\begin{smallmatrix}\alpha\\ \beta\\ \kappa\\ \gamma\end{smallmatrix}\right) \middle| \mathcal{V}\right) = \begin{cases} \inf_\eta \left\{ g_0(\alpha_0 - \eta) + \beta_0 \;\middle|\; \begin{array}{l} -\kappa_i g_i\left(\frac{\alpha_i - \eta}{-\kappa_i}\right) + \beta_i \leq 0, \;\; i = 1,\dots,m_1 \\ B\eta = \gamma, \;\; \kappa_0 = -1, \;\; \kappa_i \leq 0, \;\; i = 1,\dots,m_1 \end{array} \right\} \\ \infty \qquad \text{otherwise,} \end{cases}$$

where $\alpha \in \mathbb{R}^{(m_1+1)n_y}$, $\beta \in \mathbb{R}^{m_1+1}$, $\kappa \in \mathbb{R}^{m_1+1}$, *and* $\gamma \in \mathbb{R}^{m_2}$.

Proof. See Appendix A. □

This theorem can be used for (8). For example, let us consider one constraint from the last set of inequalities in (8):

$$q_j \geq \delta^*\left(\left(\begin{smallmatrix} -Q_{j,\cdot} \\ -R_{j,\cdot} \\ -S_{j,\cdot} \\ -T_{j,\cdot} \end{smallmatrix}\right) \middle| \mathcal{V}\right) \quad \Leftrightarrow \quad q_j \geq \inf_\eta \left\{ \beta_0 + g_0(-\widetilde{Q}^j_{0,\cdot} - \eta) \;\middle|\; \begin{array}{l} S_{j,i} g_i\left(\frac{-\widetilde{Q}^j_{i,\cdot} - \eta}{S_{j,i}}\right) \leq R_{j,i}, \;\; i = 1,\dots,m_1 \\ B\eta = T_{j,\cdot}, \;\; S_{j,0} = 1, \;\; S_{j,i} \geq 0, \;\; i = 1,\dots,m_1 \end{array} \right\}.$$

$$\Leftrightarrow \quad \begin{cases} q_j \geq \beta_0 + g_0(-\widetilde{Q}^j_{0,\cdot} - \eta) \\ S_{j,i} g_i\left(\frac{-\widetilde{Q}^j_{i,\cdot} - \eta}{S_{j,i}}\right) \leq R_{j,i} & i = 1,\dots,m_1 \\ B\eta = T_{j,\cdot}, \;\; S_{j,0} = 1, \;\; S_{j,i} \geq 0 & i = 1,\dots,m_1, \end{cases} \tag{9}$$

where $\widetilde{Q}^j \in \mathbb{R}^{(m_1+1)\times n_y}$ is a reshaped matrix from the vector $Q_{j,\cdot}$, such that the i-th row of $\widetilde{Q}^j$, i.e., $\widetilde{Q}^j_{i,\cdot} \in \mathbb{R}^p$, consists the $[(i-1)n_y + 1]$-th to in_y-th elements of $Q_{j,\cdot}$. In the first inquality of (9), we left out the "inf" and made η an extra optimization variable.

Note that in (9) only the perspectives of g_i play a role, and not the conjugates of g_i. For example, if g_i, $i = 0,\dots,m_1$, are quadratic functions, then from Theorem 2, we know that the tractable robust counterpart (9) is conic quadratic representable (see also Example 1). Finally, we note that if Fourier-Motzkin elimination would have been used to solve the dualized problem, then finally one has to solve a static robust optimization problem, for which also the support function of Theorem 2 can be used.

2.4. Extension to other two-stage robust nonlinear models

We give an example of two-stage robust nonlinear model formats where Theorem 1 can be applied. Moreover, we discuss how one can obtain the format of (1) by introducing auxiliary variables if the problem was not originally in the that format.

Example 1 (Two-stage SOC robust model). Consider a two-stage model in the form of (1) with second-order cone constraints, that is:

$$g_i(x,\zeta,y) = \left\| \tilde{A}^i(\zeta)x + \tilde{B}^i y - \tilde{b}^i(\zeta) \right\| + (h^i)^\top y \qquad i = 0,\dots,m_1, \tag{10}$$

in (1), where $\tilde{B}^i \in \mathbb{R}^{m_2 \times n_y}$, $h^i \in \mathbb{R}^{n_y}$, $\tilde{A}^i(\zeta)$ and $\tilde{b}^i(\zeta)$ as in (2) for $i = 0,\dots,m_1$. Model (1) with g_i as in (10) is not yet in the format of (1). However, if we introduce extra adjustable variables $r_i \in \mathbb{R}^{m_1}$, $i = 0,\dots,m_1$:

$$\begin{cases} g_i(x,\zeta,y,r) = \|r_i\| + (h^i)^\top y & i = 0,\dots,m_1 \\ r_i = \tilde{A}^i(\zeta)x + \tilde{B}^i y - \tilde{b}^i(\zeta) & i = 0,\dots,m_1, \end{cases} \tag{11}$$

the model (1) with g_i as in (11) can be written as:

$$\inf_{x\in\mathcal{X}} \sup_{\zeta\in\mathcal{U}} \inf_{r,y,z} \left\{ f_0(x) + \|r_0\| + (h^0)^\top y \;\middle|\; \begin{array}{ll} \zeta^\top F_{i,\cdot}(x) + f_i(x) + \|r_i\| + (h^i)^\top y \le 0 & i = 1,\dots,m_1 \\ r_i = \tilde{A}^i(\zeta)x + \tilde{B}^i y - \tilde{b}^i(\zeta) & i = 0,\dots,m_1 \\ A(\zeta)x + By = b(\zeta) & \end{array} \right\}, \tag{12}$$

which is clearly an instance of (1). If Assumption 1 is satisfied, we can apply Theorem 1 to derive the dualized model. This dualized model can then be solved with Fourier-Motzkin elimination and linear decision rules, for which case we can immediately use Theorem 2 to derive the conic quadratic representable robust counterpart. □

Example 1 shows that sometimes one has to introduce auxiliary adjustable variables in order to obtain a model that fits the format of model (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 (2) and $g : \mathbb{R}^{m_2} \to \mathbb{R}$ is a proper closed convex function, then a constraint of the form

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

can be replaced by the following system of inequalities:

$$\begin{cases} \zeta^\top F(x) + f(x) + g(r) \le 0 \\ \tilde{A}(\zeta)x + \tilde{B}y - \tilde{b}(\zeta) = r, \end{cases}$$

where $y \in \mathbb{R}^{n_y}$ is the original adjustable variable and $r \in \mathbb{R}^{m_2}$ is an additional adjustable variable. Another example is:

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

in which $g:\mathbb{R}^{m_3}\to\mathbb{R}$ is a proper closed nondecreasing convex function, and $h_k(\zeta,x,y)=\zeta^\top\tilde{F}_k(x)+\tilde{f}_k(x)+\tilde{g}_k(y)$ for all $k=1,\dots,m_3$. Such a constraint can be replaced by the following system of inequalities:

$$\begin{cases}\zeta^\top F(x)+f(x)+g(r)\le 0\\ \zeta^\top\tilde{F}_k(x)+\tilde{f}_k(x)+\tilde{g}_k(y)\le r_k \qquad k=1,\dots,m_3,\end{cases}$$

where $\tilde{g}_i:\mathbb{R}^{m_2}\to\mathbb{R}$ is a proper closed convex function, $k=1,\dots,m_3$, $r\in\mathbb{R}^{m_3}$ is again an additional adjustable variable. All these substitutions are made to ensure that the resulting system of inequalities fits into the format of model (1).

2.5. Challenges with generalizations of model formulation

The tractability of the reformulation of the model (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 (4), 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 are continuous. If the original model requires y 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. Integrality of x can be captured in the set $\mathcal{X}$ in (1). 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 $g_i(y)$ 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 adjustable variables as was done in Examples 1. 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 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 adjustable variables which are nonlinear in the adjustable variables. In that case, the benefits of the dual formulations are not clear, as we still have the same difficulties that also arose in (1) because of the nonlinearity in the adjustable variables.

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 (4) require maximization of a norm over $\mathcal{V}$ which is very difficult.

3. Bounds on the optimal value

The dualized model (4) is linear in the adjustable variables, so good solutions can be found using methods such as linear decision rules, possibly combined with Fourier-Motzkin elimination for a subset of the adjustable variables. 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 (1) to assess the quality of the solutions. Many of the ideas in this section are generalizations of the lower bound techniques discussed in Zhen et al. (2017a).

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 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:

$$\begin{aligned}
\inf_{\substack{\tau, x \in \mathcal{X} \\ y^1, \ldots, y^K}} \quad & \tau \\
\text{s.t.} \quad & f_0(x) + g_0(y^k) \leq \tau \qquad \forall k = 1, \ldots, K \\
& (\zeta^k)^\top F_{i,\cdot}(x) + f_i(x) + g_i(y^k) \leq 0 \qquad \forall i = 1, \ldots, m_1,\ k = 1, \ldots, K \\
& A(\zeta^k) + By^k = b(\zeta^k) \qquad \forall k = 1, \ldots, K.
\end{aligned} \tag{13}$$

Model (13) 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 (4). 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:

$$\begin{aligned}
\inf_{\substack{\tau, x \in \mathcal{X} \\ \lambda^1, \ldots, \lambda^K \geq 0}} \quad & \tau \\
\text{s.t.} \quad & f_0(x) + \sum_{i=1}^{m_1} v_i^k f_i(x) + d^\top \lambda^k + (w^k)^\top \left(A^0 x - b^0\right) - \sum_{i=0}^{m_1} z_i^k \leq \tau \qquad \forall k = 1, \ldots, K \\
& \sum_{j=1}^{p} D_{j,l} \lambda_j^k \geq (w^k)^\top (A^l x - b^l) + \sum_{i=1}^{m^1} (v_i^k) F_{i,l}(x) \qquad \forall l = 1, \ldots, n_\zeta, \; k = 1, \ldots, K,
\end{aligned} \tag{14}$$

which is again a standard convex optimization model.

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. *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* (13) *is also optimal for model* (1) *and their optimal objective values coincide.*

Proof. Let $\bar{\tau}, \bar{x}, \bar{y}^1, \ldots, \bar{y}^K$ be the optimal solution of (13). We know that the optimal value $\bar{\tau}$ of the sampled model (13) gives a lower bound of (1), so it is sufficient to show that $\bar{x}$ is feasible and we can construct a feasible decision rule y that gives an objective value of at most $\bar{\tau}$. Let $\zeta \in \mathcal{U}$ and write it as the convex combination of the extreme points of $\mathcal{U}$:

$$\zeta = \sum_{k=1}^{K} \alpha_k \zeta^k \tag{15}$$

for some $\alpha_1, \ldots, \alpha_K \in [0, 1]$, $\sum_{k=1}^{K} \alpha_k = 1$. We take for the adjustable variable y the following value

$$y = \sum_{k=1}^{K} \alpha_k \bar{y}^k, \tag{16}$$

with $\alpha_1, \ldots, \alpha_K$ the same values as those in the convex combination of (15). Then we have:

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

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 (13). Analogously, we can show that for $\bar{x}$ and decision rule y from (16) we have $f_0(x) + g_0(y) \leq \bar{\tau}$ for all $\zeta \in \mathcal{U}$. Hence, the optimal objective value of (1) is at most $\bar{\tau}$. □

Of course, the set of extreme points of a polyhedral uncertainty set $\mathcal{U}$ is in practice way too large. As demonstrated 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 adjustable variables.

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 (14):

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \sup_{1 \leq k \leq K} \; f_0(x) + \sum_{i=1}^{m_1} v_i^k \left(\zeta^\top F_{i,\cdot}(x) + f_i(x)\right) + \left(A(\zeta)x - b(\zeta)\right)^\top w^k - \sum_{i=0}^{m_1} z_i^k, \tag{17}$$

which is similar to (6), but with the inner supremum 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 (17):

$$\zeta^k \in \arg\max_{\zeta \in \mathcal{U}} \left\{ \sum_{i=1}^{m_1} v_i^k \left(\zeta^\top F_{i,\cdot}(x)\right) + \left(w^k\right)^\top \left(A(\zeta)x - b(\zeta)\right) \right\}. \tag{18}$$

The resulting set of scenarios $\{\zeta^1, \ldots, \zeta^K\}$ can then be used in the sampled model (13). One can now solve either the primal sampled model (13), which is a convex optimization model, or the dual sampled model (14), 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 (13) or (14) 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 (1), and the scenarios have been obtained by (18), then we can show that the lower bound from (13) is always higher (or equal to) (14). 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 1 (Right-hand-side uncertainty). Model (1) has right-hand-side uncertainty if $F_{i,\cdot} = 0$ for all $i = 1,\dots,m_1$ 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 (18) independent of the here-and-now decision x as the only terms depending on ζ are $(w^k)^\top b(\zeta)$.

Theorem 4. *Let* $\{(u^1, w^1, v^1, z^1), \dots, (u^K, w^K, v^K, z^K)\}$ *be a finite set of dual scenarios and* $\{\zeta^1, \dots, \zeta^K\}$ *be a set of primal scenarios obtained from* (18). *If there is only right-hand-side uncertainty in model* (1), *then the lower bound from* (13) *is at least as tight as the lower bound from* (14).

Proof. By duality for linear programming, (14) is equivalent to (17). The latter formulation can be written as

$$\inf_{x \in \mathcal{X}} \sup_{k \in \{1,\dots,K\}} \left\{ f_0(x) + \sum_{i=1}^{m_1} v_i^k f_i(x) + \left(w^k\right)^\top \left(\bar{A}x - b(\zeta^k)\right) - \sum_{i=0}^{m_1} z_i^k \right\}, \tag{19}$$

where ζ^k are the primal scenarios obtained by (18). Since (u^k, w^k, v^k, z^k) are in $\mathcal{V}$ for all $k = 1,\dots,K$, the value of (19) must be smaller than or equal to

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

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 (13). Hence, the optimal objective value of the model (13) is at least as high as the optimal objective value of (14). □

We emphasize that the right-hand-side uncertainty definition is stated for models that are of the format (1). This means that it could also apply to models where auxiliary adjustable variables are introduced and right-hand-side uncertainty is only visible in the final formulation of the model, e.g., (12) with $F_{i,\cdot} = 0$ for all $i = 1,\dots,m_1$, $A^i(\zeta) = \bar{A}^i$ for all $i = 0,\dots,m_1$, and $A(\zeta) = \bar{A}$.

4. Example 1: distribution on a network with commitments

4.1. Problem formulation

This problem is adapted from Bertsimas and de Ruiter (2016). 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:

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

where $\hat{\zeta}_i \in \mathbb{R}_+$ denotes the maximum demand at location i, $i = 1,...,N$, and $\Gamma \in \mathbb{R}_+$ denotes the maximum total demand. 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:

$$\inf_{x \in \mathcal{X}, z} \sup_{\zeta \in \mathcal{U}} \inf_{y \geq 0} \left\{ \sum_{i=1}^N c_i x_i + \sum_{i,j=1}^N t_{ij} y_{ij} + \frac{1}{2} \sum_{i,j=1}^N t_{ij} (y_{ij} - z_{ij})^2 \ \middle| \ \sum_{j=1}^N y_{ji} - \sum_{j=1}^N y_{ij} \geq \zeta_i - x_i \quad i = 1,...,N \right\}, \tag{20}$$

where the third term in the objective of (20) captures the cost of contract violation, and $\mathcal{X} = \left\{ x \in \mathbb{R}_+^N \mid e^\top x \geq \Gamma, \ x_i \leq K_i \quad i = 1,...,N \right\}$. The constraints in (20) are 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 $\mathcal{X}$ restrict the capacity of the stock at location i to at most K_i, $i = 1,...,N$, as well as the total stock to be at least the maximum demand. The dualized formulation we obtain after consecutive dualization over the adjustable variables y and the uncertain parameters ζ is given below:

$$\inf_{\substack{x \in \mathcal{X} \\ z,\tau}} \sup_{(u,v,w) \in \mathcal{V}} \inf_{\lambda \geq 0} \left\{ \sum_{i=1}^N c_i x_i + \tau \ \middle| \ \begin{array}{l} \sum_{i=1}^N (\hat{\zeta}\lambda_i - u_i x_i) + \Gamma \lambda_0 - \sum_{i,j=1}^N \left[(u_j - u_i - t_{ij} - v_{ij}) z_{ij} + \frac{1}{2} w_{ij} \right] \leq \tau \\ \lambda_0 + \lambda_i \geq u_i \qquad i = 1,...,N \end{array} \right\}, \tag{21}$$

where $\mathcal{V} = \{(u,v,w) \geq 0 : \ (u_i - u_j + v_{ij} - t_{ij})^2 \leq w_{ij} t_{ij} \quad \forall i,j = 1,...N\}$. Now we can apply Fourier-Motzkin elimination and linear decision rules to solve (21).

4.2. 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 $\hat{\zeta}$ 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 1 depicts a distribution on a network obtained from solving (21) 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`).

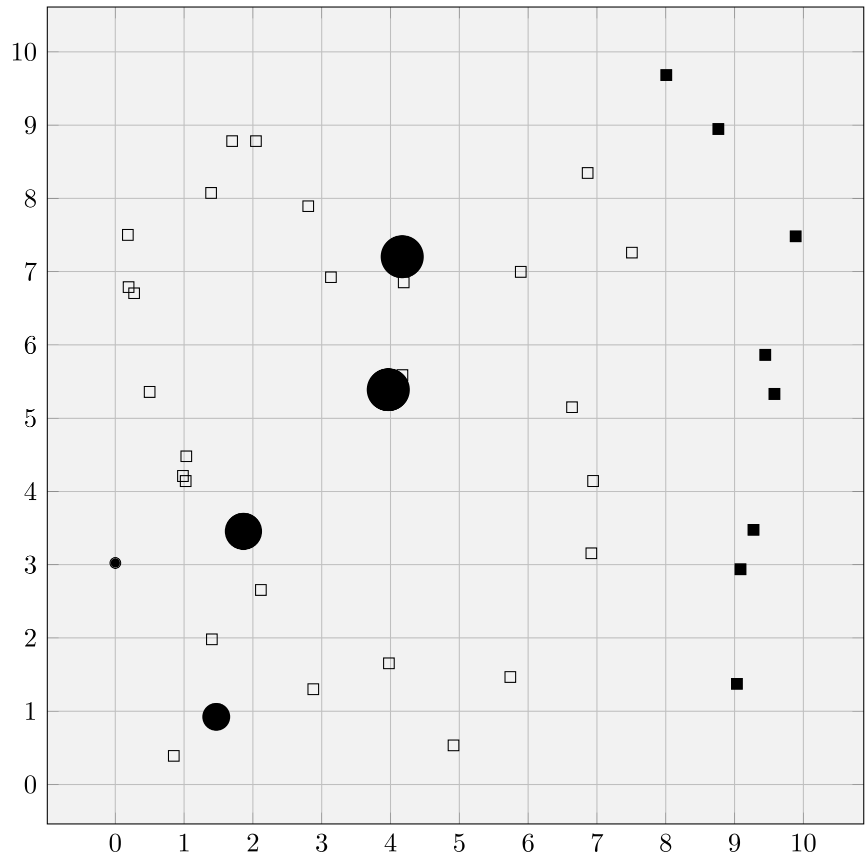

Figure 1 Stock allocation for $N = 40$ with 35 stores (squares) and 5 warehouses (circles) for one random instance. The filled dots have stock and the larger the dots are, the more stock is allocated.

4.3. Results

We first consider a small instance and present the results in Table 1. One can observe that the solutions converge to optimality as more adjustable variables in (21) 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 (20) because the adjustable variables appear nonlinearly in the model. By solving (20) and (21) 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 (20) and (21) 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 (20) 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 4). Hence, we only focus on the LB-P lower bounds for the rest of this chapter.

Table 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 (20) and (21) with static decision rules, respectively. D-L is obtained from solving (21) with linear decision rules. LB-P and LB-D denote the lower bounds obtained from the primal scenarios (see §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 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), 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. 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.

5. Example 2: equilibrium of system with piecewise-linear springs

5.1. Problem formulation

The problem described in this section is adopted from Lobo et al. (1998). We consider a mechanical system that consists of N nodes at positions $x_1,\dots,x_N \in \mathbb{R}^2$, with node i connected to node $i+1$, for $i=1,\dots,N-1$, by a nonlinear spring. The nodes x_1 and x_N are fixed at given values a and b,

Table 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 (20) with static decision rules. D-L is obtained from solving (21) with linear decision rules. LB-P denotes the lower bounds obtained from the primal scenarios (see §3.3). Time(s) reports the computation time (in seconds) for solving D-L. * means the computation time needed exceeds 1 hour. N.A. represents not applicable. 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 Lot-sizing problem with $N \in \{30, 40, 50, 60\}$. P-S is obtained from solving (20) with static decision rules. D-L is obtained from solving (21) with linear decision rules. LB-P denotes the lower bounds obtained from the primal scenarios (see §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

respectively. The tension in spring i is a nonlinear function of the distance between its endpoints, i.e., $\|x_i - x_{i+1}\|_2$:

$$s\left(\|x_i - x_{i+1}\|_2 - l_i^0\right)_+,$$

where $z_+ = \sup\{z, 0\}$, $s \in \mathbb{R}_+$ is the stiffness of the springs, and $l_i^0 \in \mathbb{R}_+$ is the natural (no tension) length of spring i. In this model the springs can only produce positive tension (which would be the case if they buckled under compression). Each node has a mass of weight w attached to it. This is shown in Figure 2. The problem is to compute the equilibrium configuration of the system, i.e., values of $x_1, \ldots, x_N$ such that the net force on each node is zero. This can be done by finding the minimum energy configuration, i.e., solving a second-order cone optimization problem:

$$\begin{aligned} \inf_{x \geq 0} \; & w \sum_{i=1}^{N} x_{i2} + \frac{s}{2} \sum_{i=1}^{N-1} \left[\left(\|x_i - x_{i+1}\|_2 - l_i^0\right)_+\right]^2 \\ \text{s.t.} \; & x_1 = a, \quad x_N = b, \end{aligned} \tag{22}$$

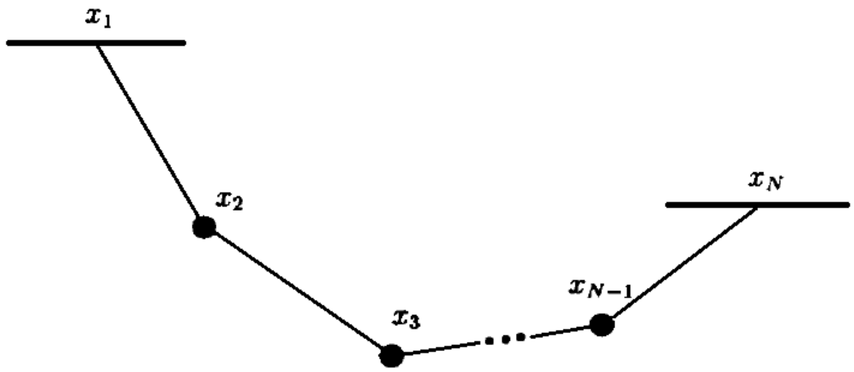

Figure 2 System of nodes (weights) connected by springs from Lobo et al. (1998). The first and last node positions, i.e., x_1 and x_N, are fixed.

where x_{i2} is the second element of the vector x_i. For more detailed description of this problem, we refer to the original paper Lobo et al. (1998). Suppose the length of the springs are uncertain. The uncertainty may arise due to variations in the production process Of course other parameters, e.g., weight(w), stiffness(s), initial location of x_1 and x_N, may also be uncertain. Here we focus on uncertainty in the length of the springs, i.e., $l(\zeta) = l^0 - \zeta$ (because only positive tension is considered), and the uncertain parameter $\zeta \in \mathbb{R}^{N-1}$ resides in a budget uncertainty set:

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

where $\hat{\zeta}_i \in \mathbb{R}_+$ denotes the maximum deviation from the nominal length l^0 of spring i, $i = 1, ..., N-1$, and $\Gamma \in \mathbb{R}_+$ denotes the maximum total deviation of the springs. The minimum energy configuration model becomes a robust optimization model:

$$\inf_{x \geq 0} \sup_{\zeta \in \mathcal{U}} \left\{ w \sum_{i=1}^{N} x_{i2} + \frac{s}{2} \sum_{i=1}^{N-1} \left[\left(\|x_i - x_{i+1}\|_2 - l_i(\zeta) \right)_+ \right]^2 \ \middle| \ x_1 = a, \quad x_N = b \right\}, \tag{23}$$

which can be rewritten as a two-stage robust optimization problem:

$$\inf_{x \geq 0} \sup_{\zeta \in \mathcal{U}} \inf_{y \geq 0} \left\{ w \sum_{i=1}^{N} x_{i2} + \frac{s}{2} \sum_{i=1}^{N-1} y_i^2 \ \middle| \ \begin{array}{l} \|x_i - x_{i+1}\|_2 - l_i(\zeta) \leq y_i \ \ i = 1, ..., N-1 \\ x_1 = a, \quad x_N = b \end{array} \right\}. \tag{24}$$

It can be verified that models (23) and (24) are equivalent, that is, eliminating all the y_i's for $i = 1, \ldots, N-1$ in (24) we obtain (23). We solve the dualized formulation of (24) via linear decision rules. Note that here the strong relatively complete recourse assumption is satisfied.

5.2. Numerical setting

We consider $N \in \{15, 20, 30, 45, 60, 100\}$ nodes that are connecting $N-1$ springs. The nodes x_1 and x_N are fixed at given values $a = (0, 90)$ and $b = (100, 50)$, respectively. The natural (no tension) nominal length is $l_i^0 = 1 + \epsilon_i$, where ϵ_i is a random number drawn from a uniform distribution $U(0,4)$, $i = 1, \ldots, N-1$, and the stiffness of the springs is $s = 2$. Each node has a mass of weight $w = \frac{1}{10}$ attached to it. The upper-bound $\hat{\zeta}_i$ is set at $15\% l_i^0$ for $i = 1, \ldots, N-1$, and $\Gamma = \frac{1}{2} e^\top \hat{\zeta}$. The computations is carried out with MOSEK 8.1 (MOSEK ApS, 2017) on an Intel(R) Xeon(R) E3-1241 v3 Windows computer running at 3.50GHz with 16GB of RAM. All modeling was done using the modeling package XProg (`http://xprog.weebly.com`).

5.3. Results

Figure 3 illustrates the static and robust locations of the nodes for $N = 45$, which shows that in order to minimize energy configuration under length uncertainty, in the solution from linear decision rules consecutive nodes are placed closer to each other than in the solution from static decision rules. Figure 4 depicts the robust locations obtained from solving the dualized model of (24) with linear decision rules. It shows that as N increases, the curvature of the connection between x_1 and x_N becomes severer; if N is large enough, i.e., $N = 100$, then there are too many nodes with positive weights, all the useless nodes will simply be closely placed on the ground.

Table 4 shows that the approximations from solving the dualized model of (24) via linear decision rules are tight, because the corresponding objective values, i.e., D-L, equal to the lower bounding objective values, i.e., LB-P. For small N, we observe that the P-S values are larger than the D-L values, which means that the approximated solutions obtained via static decision rules are suboptimal. Since the robust problem (24) becomes easier to solve as N becomes larger, the objective values P-S and D-L becomes close. P-N gives the optimal objective values of the nominal problem, which can be seen as lower bounds of P-S and D-L.

Table 4 Equilibrium of system with $N-1$ piecewise-linear springs for $N \in \{15, 20, 30, 45, 65, 100\}$. P-S is obtained from solving (24) with static decision rules. D-L is obtained from solving the dualized model of (24) with linear decision rules. LB-P denotes the lower bounds obtained from the primal scenarios (see §3.3). Time(s) reports the computation time (in seconds) for solving D-L.

N	15	20	30	45	65	100
P-S	535.8	344.5	254.1	213.7	189.1	180.7
D-L	507.6	320.2	239.0	202.8	183.5	180.7
LB-P	507.6	320.2	239.0	202.8	183.5	180.7
Time(s)	0.05	0.06	0.13	0.42	1.50	4.97

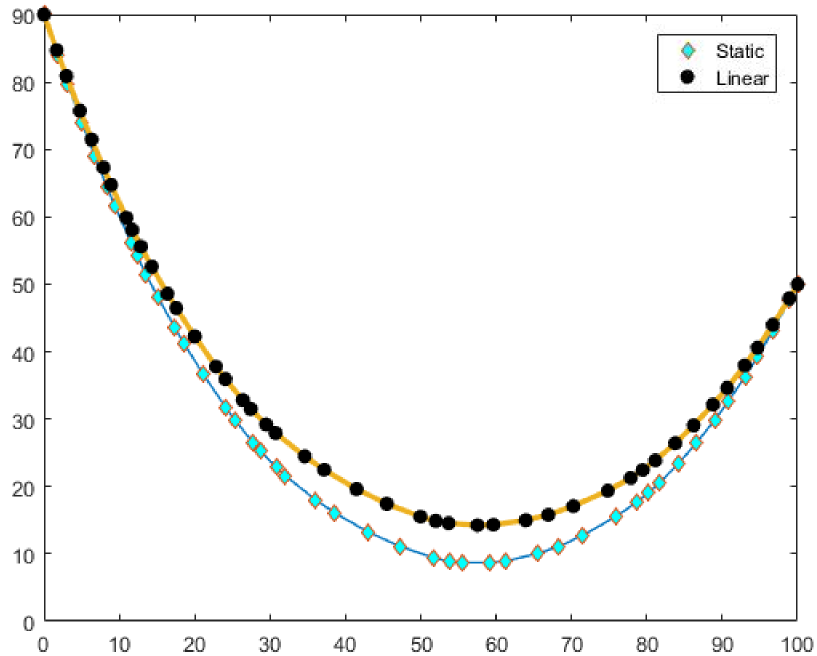

Figure 3 System of nodes (weights) connected by 44 springs for $N = 45$. The diamonds and dots represent the robust locations of the nodes from solving (24) and its dualized formulation with static decision rules and linear decision rules, respectively.

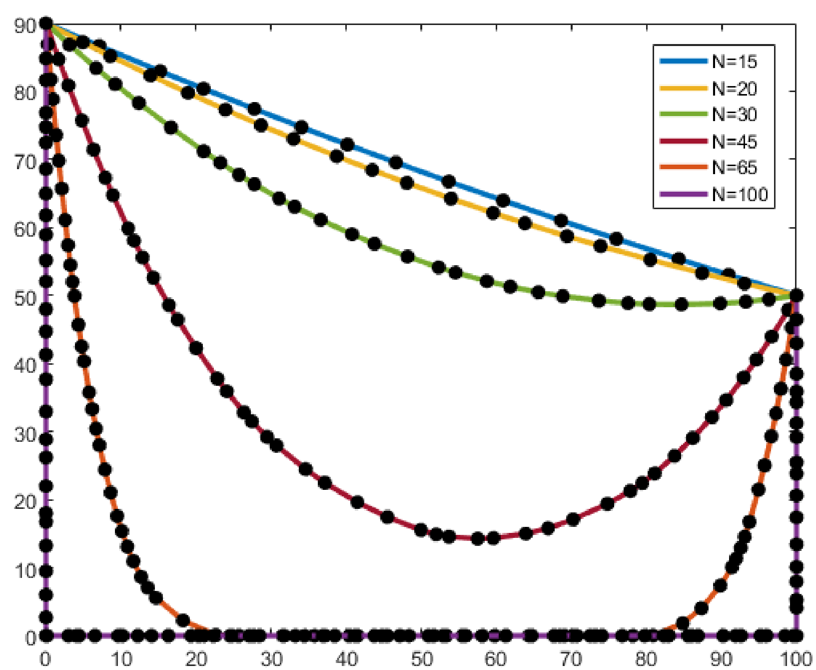

Figure 4 System of nodes (weights) connected by $N-1$ springs for $N \in \{15, 20, 30, 45, 65, 100\}$. The dots represent the robust location of the nodes.

6. Conclusion and future research

In this paper we reformulate the two-stage robust nonlinear problem with fixed recourse and a polyhedral uncertainty set into an equivalent two-stage robust linear problem, for which approaches as Fourier-Motzkin elimination and (non)linear decision rules can be used. We also describe how to effectively obtain lower bounds on the optimal objective value by linking the realizations in the new dualized uncertainty set to realizations in the original uncertainty set. Two numerical examples show the effectiveness and applicability of the new approach.

The dualized models with linear decision rules admits tractable robust counterparts, whereas, the original models do not. One interesting future research direction would be to search for the decision rules for the original two-stage nonlinear models that correspond to the linear decision rules for the dualized model.

Another interesting direction would be to extend our approach to two-stage distributionally robust nonlinear optimization models with affine moment functions and polyhedral support.

Acknowledgments

The research of the first author is partially supported by the Netherlands Organisation for Scientific Research (NWO) Talent Grant 406-14-067. The second author is supported by the NWO Grant 613.001.208.

A. Appendix: Proof of Theorem 2

First of all, let us represent $\mathcal{V}$ as the intersection of m_1+2 sets:

$$\mathcal{V} = \left(\bigcap_{i=0}^{m_1} \mathcal{V}_i \right) \bigcap \mathcal{W},$$

where

$$\begin{aligned}
\mathcal{V}_0 &= \{(u_0, v_0, z_0) \mid v_0 = 1,\ g_0^*(u_0) \le z_0\} \\
\mathcal{V}_i &= \left\{(u_i, v_i, z_i) \;\middle|\; v_i \ge 0,\ v_i g_i^*(\tfrac{u_i}{v_i}) \le z_i \right\} \quad i = 1, \ldots, m_1 \\
\mathcal{W} &= \left\{(u_1, \ldots, u_{m_1}, w) \;\middle|\; \sum_{i=0}^{m_1} u_i = -B^\top w \right\}.
\end{aligned}$$

Suppose the sets $\mathcal{V}_i$, $i = 0, \ldots, m_1$, and $\mathcal{W}$ satisfy the Slater condition, we may apply Rockafellar (1970, Corollary 16.4.1), which gives:

$$\delta^*\left(\left(\begin{smallmatrix}\alpha\\ \beta\\ \kappa\\ \gamma\end{smallmatrix}\right) \;\middle|\; \mathcal{V} \right) = \inf_{\mu,\theta} \left\{ \sum_{i=0}^{m_1} \delta^*\left(\left(\begin{smallmatrix}\mu_i\\ \beta_i\\ \kappa_i\end{smallmatrix}\right) \;\middle|\; \mathcal{V}_i \right) + \delta^*\left(\left(\begin{smallmatrix}\theta_0\\ \vdots\\ \theta_{m_1}\\ \gamma\end{smallmatrix}\right) \;\middle|\; \mathcal{W} \right) \;\middle|\; \mu_i + \theta_i = \alpha_i,\ i = 0, \ldots, m_1 \right\}. \tag{25}$$

We can now compute the conjugate of the support function for $\mathcal{V}_i$, $i = 0, \ldots, m_1$, and $\mathcal{W}$ separately, and then substitute the obtained conjugates into (25). The conjugate of the support function for $\mathcal{V}_i$, $i = 1, \ldots, m_1$, can be rewritten as follows (Ben-Tal et al., 2015, Lemma 9):

$$\delta^*\left(\left(\begin{smallmatrix}\mu_i\\ \beta_i\\ \kappa_i\end{smallmatrix}\right) \;\middle|\; \mathcal{V}_i \right) = \inf_{\omega_i \ge 0} \left\{ \omega_i h^*\left(\frac{\mu_i}{\omega_i}, \frac{\beta_i}{\omega_i}, \frac{\kappa_i}{\omega_i} \right) \right\}$$

where h^* is the conjugate of the function h, which is defined as:

$$h(u_i, v_i, z_i) = v_i g_i^* \left(\frac{u_i}{v_i} \right) - z_i.$$

By definition, we have:

$$\begin{aligned}
h^*(\mu_i, \beta_i, \kappa_i) &= \sup_{v_i \ge 0, u_i, z_i} \left\{ u_i^\top \mu_i + v_i \beta_i + z_i \kappa_i - (v_i g_i)^*(u_i) + z_i \right\} \\
&= \sup_{v_i \ge 0} \{ v_i \beta_i + v_i g_i(\mu_i) \} + \begin{cases} 0 & \text{if } \kappa_i = -1 \\ \infty & \text{otherwise} \end{cases} \\
&= \begin{cases} 0 & \text{if } g_i(\mu_i) + \beta_i \le 0 \text{ and } \kappa_i = -1 \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Therefore, we have

$$\delta^*\left(\left(\begin{smallmatrix}\mu_i\\ \beta_i\\ \kappa_i\end{smallmatrix}\right) \;\middle|\; \mathcal{V}_i \right) = \begin{cases} 0 & \text{if } \exists \omega_i \ge 0 : \omega_i g_i\left(\frac{\mu_i}{\omega_i}\right) + \beta_i \le 0 \text{ and } \kappa_i = -\omega_i \\ \infty & \text{otherwise.} \end{cases} \tag{26}$$

In a similar way, we obtain the conjugate of the support function for $\mathcal{V}_0$:

$$\delta^*\left(\left(\begin{smallmatrix}\mu_i\\ \beta_i\\ \kappa_i\end{smallmatrix}\right)\ \middle|\ \mathcal{V}_0\right)=\begin{cases}\inf_{\mu_0}\ g_0(\mu_0)+\beta_0 & \text{if } \kappa_0=-1\\ \infty & \text{otherwise.}\end{cases} \tag{27}$$

We use LP strong duality to compute the conjugate of the support function for $\mathcal{W}(\neq\emptyset)$:

$$\delta^*\left(\left(\begin{smallmatrix}\theta_0\\ \vdots\\ \theta_{m_1}\\ \gamma\end{smallmatrix}\right)\ \middle|\ \mathcal{W}\right)=\inf_{\eta}\left\{0\ \middle|\ \eta=\theta_0=\cdots=\theta_{m_1},\ B\eta=\gamma\right\}. \tag{28}$$

Finally, by substituting (26), (27) and (28) into (25) yields:

$$\begin{aligned}\delta^*\left(\left(\begin{smallmatrix}\alpha\\ \beta\\ \kappa\\ \gamma\end{smallmatrix}\right)\ \middle|\ \mathcal{V}\right)&=\begin{cases}\inf_{\mu_0}\ g_0(\mu_0)+\beta_0 & \text{if}\ \ \exists\omega\geq 0,\eta,\mu,\theta:\ \begin{cases}\mu_i+\theta_i=\alpha_i, & i=0,\ \ldots,m_1\\ \omega_i g_i\left(\frac{\mu_i}{\omega_i}\right)+\beta_i\leq 0, & i=1,\ \ldots,m_1\\ \kappa_0=-1,\ \kappa_i=-\omega_i, & i=1,\ \ldots,m_1\\ \eta=\theta_0=\cdots=\theta_{m_1}, B\eta=\gamma & \end{cases}\\ \infty & \text{otherwise}\end{cases}\\ &=\begin{cases}\inf_{\eta}\ \left\{g_0(\alpha_0-\eta)+\beta_0\ \middle|\ \begin{array}{c}-\kappa_i g_i\left(\frac{\alpha_i-\eta}{-\kappa_i}\right)+\beta_i\leq 0,\quad i=1,\ \ldots,m_1\\ B\eta=\gamma,\ \kappa_0=-1,\ \kappa_i\leq 0,\ i=1,\ \ldots,m_1\end{array}\right\}\\ \infty & \text{otherwise,}\end{cases}\end{aligned}$$

where the last equality follows from the substitutions $\kappa_i=-\omega_i$ and $\eta=\theta_0=\cdots=\theta_{m_1}$, for $i=1,\ \ldots,m_1$, and $\alpha_i=\mu_i+\theta_i$, for $i=0,\ \ldots,m_1$. $\square$

References

A. Ardestani-Jaafari and E. Delage. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494, 2016.

A. Ardestani-Jaafari and E. Delage. The value of flexibility in robust location–transportation problems. *Transportation Science*, pages 1–21, 2017. doi: 10.1287/trsc.2016.0728. Published online in Articles in Advance.

A. Atamtürk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, 2007.

A. Bemporad, F. Borrelli, and M. Morari. Min-max control of constrained uncertain discrete-time linear systems. *IEEE Transactions Automatic Control*, 48(9):1600–1606, 2003.

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

A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

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

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

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

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

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

D. Bertsimas and H. Bidkhori. On the performance of affine policies for two-stage adaptive optimization: a geometric perspective. *Mathematical Programming*, 153(2):577–594, 2015.

D. Bertsimas and F. de Ruiter. Duality in two-stage adaptive linear optimization: faster computation and stronger bounds. *INFORMS Journal on Computing*, 28(3):500–511, 2016.

D. Bertsimas and V. Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, 134(2):491–531, 2012.

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

D. Bertsimas, D. Iancu, and P. Parrilo. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

D. Bertsimas, E. Litvinov, X. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, 2013.

J. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer Science & Business Media, 2011.

O. Boni and A. Ben-Tal. Adjustable robust counterpart of conic quadratic problems. *Mathematical Methods of Operations Research*, 68(2):211–233, 2008.

G. Calafiore. Multi-period portfolio optimization with linear control policies. *Automatica*, 44(10):2463–2473, 2008.

G. Calafiore. An affine control method for optimal dynamic asset allocation with transaction costs. *SIAM Journal on Control and Optimization*, 48(4):2254–2274, 2009.

V. Gabrel, M. Lacroix, C. Murat, and N. Remli. Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics*, 164(1):100–11, 2014.

B. Gorissen and D. den Hertog. Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–43, 2013.

E. Guslitzer. Uncertainty-immunized solutions in linear programming. Master's thesis, Technion-Israel Institute of Technology, 2002. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.623.2382&rep=rep1&type=pdf`.

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

D. Iancu, M. Sharma, and M. Sviridenko. Supermodularity and affine policies in dynamic robust optimization. *Operations Research*, 61(4):941–956, 2013.

M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1-3):193–228, 1998.

MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1*, 2017. URL `http://docs.mosek.com/8.1/toolbox.pdf`.

T. Ng and C. Sy. An affine adjustable robust model for generation and transmission network planning. *International Journal of Electrical Power & Energy Systems*, 60:141–152, 2014.

F. Ordóñez and J. Zhao. Robust capacity expansion of network flows. *Networks*, 50(2):136–145, 2007.

M. Pınar and R. H Tütüncü. Robust profit opportunities in risky financial portfolios. *Operations Research Letters*, 33(4):331–340, 2005.

P. Rocha and D. Kuhn. Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *European Journal of Operational Research*, 216(2):397–408, 2012.

R. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

K. Roos, M. Balvert, B. Gorissen, and D. den Hertog. A universal and structured way to derive dual optimization problem formulations. *Optimization Online*, 2016. URL http://www.optimization-online.org/DB_HTML/2016/09/5638.html.

A. Takeda, S. Taguchi, and R. Tütüncü. Adjustable robust optimization models for a nonlinear two-period system. *Journal of Optimization Theory and Applications*, 136(2):275–295, 2008.

W. Wiesemann, D. Kuhn, and B. Rustem. Robust resource allocations in temporal networks. *Mathematical Programming*, 135(1):437–471, 2012.

J. Zhen and D. den Hertog. Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS Journal on Computing*, 30(1):31–42, 2017.

J. Zhen, F.J.C.T. de Ruiter, and D. den Hertog. Robust optimization for models with uncertain SOC and SDP constraints. *Optimization Online*, 2017a. URL http://www.optimization-online.org/DB_FILE/2017/12/6371.pdf.

J. Zhen, D. den Hertog, and M. Sim. Adjustable robust optimization via Fourier-Motzkin elimination. *Operations Research*, to appear, 2017b.