

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 minimization problems 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, which implies that the adjustable variables should appear linearly in the objective and constraints. 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 all existing approaches for adjustable robust linear problems 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. We show how to recover the linear decision rule to the original primal problem and extend our method to multistage cases. This paper also describes how to effectively obtain lower bounds (for minimization problems) 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;

* Corresponding author: jianzhe.zhen@epfl.ch

Ben-Tal et al., 2015; Bertsimas and Sim, 2004). In robust optimization one seeks for a fixed solution that minimizes the worst-case cost, and is immunized against any realization of the uncertain parameter in the uncertainty set. 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 (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, 2018; 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, solutions can be efficiently approximated using linear decision rules. Rather than allowing the adjustable variable to depend arbitrarily on the uncertain parameter, affine decision rules restrict the dependence to be linear. 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 can easily incorporate nonanticipative constraints, and the resulting tractable counterpart is of the same complexity class as the static robust version where all decisions have to be made here-and-now. For some special classes of two-stage robust linear problems, static solutions are shown to be optimal (Bertsimas et al., 2015; Marandi and den Hertog, 2018). The optimality of linear decision rules for some special multistage problems has been established by (Bertsimas et al., 2010; Iancu et al., 2013), while there are other papers that establish optimality or provide theoretical a-priori bounds on the objective value (Bertsimas and Bidkhori, 2015; Bertsimas and Goyal, 2012). For more general settings, Hanasusanto and Kuhn (2018) and Xu and Burer (2018) propose independently equivalent copositive programming reformulations for two-stage robust linear optimization problems and develop conservative semidefinite approximations for the reformulations, which has been recently extended to multistage cases with right-hand-side uncertainty (Hanasusanto and Xu, 2018). Another recently developed method is Fourier-Motzkin elimination for adjustable robust optimization. This method can solve adjustable robust linear optimization models with a small

number of adjustable variables and with general convex uncertainty sets to optimality by eliminating the adjustable variables. For larger problems one 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. Pinar 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. This

significantly extends the approach of Bertsimas and de Ruiter (2016) where only linear problems are considered. Note that in the linear case studied by Bertsimas and de Ruiter (2016), the original adjustable robust optimization models could be solved with techniques as Fourier-Motzkin elimination, linear and nonlinear decision rules, Benders decomposition, and the column-and-constraint generation (CCG) method of Zeng and Zhao (2013) as well, which is not the case (at least, not directly) for the nonlinear problems considered in this paper. However, in this paper we show that the dual of the nonlinear problem is linear in the adjustable variables, and hence for this dual problem the above mentioned well-known adjustable linear robust optimization techniques can be used. Moreover, and maybe surprisingly, by applying a new relaxation technique, we are also able to establish a non-trivial relation between linear decision rules for the original nonlinear problem and its equivalent linear reformulation, which enables us to extend our approach to solve multistage robust nonlinear optimization models.

Lastly, we show how scenarios in the uncertainty set of the primal and dualized version are tied to each other. We propose a new lower bounding scheme that also significantly improves the method by Bertsimas and de Ruiter (2016) for the linear case. To summarize, our contributions in this paper 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 can again be interpreted as a two-stage robust problem but now with adjustable variables that appear linearly. We show that this formulation 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 linear optimization in the literature can be used to find solutions.

2. We apply a new relaxation technique to establish a close relation between linear decision rules for the original nonlinear problem and its equivalent dual (linear) reformulation. This relation enables us to extend our approach to solve multistage robust optimization problems.

3. 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.

4. We show that we can use our method to efficiently solve practical adjustable 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 paper is organized as follows. In §2 we present our framework and derive our dualized formulation. We recover the decision rule for the original primal problem and extend our method to multistage cases. 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 $g : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}$ and is defined by:

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

where $\text{dom}(g)$ 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:

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

The perspective $h : \mathbb{R}^{n_\nu} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ of a closed convex function $f : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}$ 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 $h(\nu, 0) = \delta_{\text{dom } f^*}^*(\nu)$. For ease of exposition, we use $tf(\nu/t)$ to denote the perspective function $h(\nu, t)$ for the rest of this paper.

2. Linear dual formulation

2.1. Framework

We consider the following general two-stage robust nonlinear minimization problem:

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \inf_y \left\{ f_0(x) + g_0(y) \mid \zeta^\top F_i(\cdot)(x) + f_i(x) + g_i(y) \leq 0, i = 1, \dots, m_1, A(x)\zeta + By = b(x) \right\}, \quad (1)$$

where $\mathcal{X} \subseteq \mathbb{R}^{n_x}$, the functions $f_i : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^{n_y} \rightarrow \mathbb{R}$ are convex for all $i = 0, \dots, m_1$, $F_i(\cdot)(x) = (F_{i1}(x), \dots, F_{in_\zeta}(x))$ and $F_{ij} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ are real valued functions for all $i = 1, \dots, m_1$, and $j = 1, \dots, n_\zeta$. The matrices $A(x) \in \mathbb{R}^{m_2 \times n_\zeta}$ and the vector $b(x) \in \mathbb{R}^{m_2}$ depend on $x \in \mathbb{R}^{n_x}$ in an affine way:

$$A(x) = A^0 + \sum_{l=1}^{n_x} A^l x_l, \quad b(x) = b^0 + \sum_{l=1}^{n_x} b^l x_l, \quad (2)$$

with $A^l \in \mathbb{R}^{m_2 \times n_\zeta}$ and $b^l \in \mathbb{R}^{m_2}$ for all $l = 0, \dots, n_x$. We consider the *fixed recourse* case, the functions g_i , $i = 0, \dots, m_1$, and the matrix B do not depend on ζ . Loosely speaking, fixed recourse implies

that there are no direct interaction terms between ζ and y , such as products $\zeta^\top y$ etc. Throughout this paper we focus on nonempty polyhedral uncertainty sets of the form:

$$\mathcal{U} = \{\zeta \geq 0 : D\zeta \leq d\}, \quad (3)$$

where $D \in \mathbb{R}^{p \times n_\zeta}$ and $d \in \mathbb{R}^p$. Without loss of generality, the uncertain parameter ζ in (3) is assumed to be nonnegativity, and in the case that ζ is a free variable, one can replace ζ with $\zeta^+ - \zeta^-$ in (1), where $\zeta^+, \zeta^- \geq 0$ (Chen et al., 2008). In the case that problem (1) has a non-polyhedral uncertainty set, we postpone the discussion to §2.6. We emphasize that the framework in (1) is flexible enough to even incorporate functions $g(x, y, \zeta)$, where x, y , and ζ are jointly convex, as shown in the following example.

EXAMPLE 1. Let us consider the following variant version of (1):

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \inf_y \left\{ f_0(x) + g_0(A_0(x)\zeta + B_0y - b_0(x)) \left| \begin{array}{l} \zeta^\top F_i(x) + f_i(x) + \dots \\ g_i(A_i(x)\zeta + B_iy - b_i(x)) \leq 0, \quad i = 1, \dots, m_1 \end{array} \right. \right\},$$

where $A_i(x)$ and $b_i(x)$ affinely depend on $x \in \mathbb{R}^{n_x}$ as defined in (2) for $i = 0, \dots, m_1$. This problem can be reformulated into the format of (1) by simply introducing some auxiliary adjustable variables:

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \inf_{y, \{r_i\}_{i=0}^{m_1}} \left\{ f_0(x) + g_0(r_0) \left| \begin{array}{l} \zeta^\top F_i(x) + f_i(x) + g_i(r_i) \leq 0, \quad i = 1, \dots, m_1 \\ A_i(x)\zeta + B_iy - b_i(x) = r_i, \quad i = 0, \dots, m_1 \end{array} \right. \right\}. \quad (4)$$

A subclass of problem (4) are problems with robust second-order cone or robust semidefinite constraints that are jointly convex in x and ζ , see Zhen et al. (2017a). Suppose now that the constraint constitutes a composition of convex functions:

$$\zeta^\top F(x) + f(x) + g\left(\zeta^\top \tilde{F}(x) + \tilde{f}(x) + \tilde{g}(y)\right) \leq 0,$$

where $g: \mathbb{R}^{m_3} \rightarrow \mathbb{R}$ is a non-decreasing convex function in each of its arguments, the functions $F: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_\zeta}$, $\tilde{F}: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_\zeta \times m_3}$, $f: \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ and $\tilde{f}: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{m_3}$ are convex in x , and $\tilde{g}: \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{m_3}$ is convex in y . Such a constraint can also be represented as a system of inequalities in the form of (1):

$$\begin{cases} \zeta^\top F(x) + f(x) + g(r) \leq 0 \\ \zeta^\top \tilde{F}(x) + \tilde{f}(x) + \tilde{g}(y) \leq r \end{cases}$$

where $r \in \mathbb{R}^{m_3}$ is an auxiliary adjustable variable. \square

To the best of our knowledge, there is no existing literature that consider two-stage robust problems that are (not necessarily jointly) convex in the decision variable x , the uncertain parameter ζ , and the adjustable variable y .

Model (1) is generally intractable. Even if we impose linear decision rules $y = y_0 + \sum_{j=1}^{n_\zeta} y_j \zeta_j$ to the wait-and-see decision variables, the resulting problem remains difficult to solve as the objective

and constraint functions contain terms $g_i(y_0 + \sum_{j=1}^{m_\zeta} y_j \zeta_j)$ for every $i = 0, \dots, m_1$, which is convex in the uncertain parameters if g_i is nonlinear and convex, and robust optimization techniques such as described in Ben-Tal et al. (2015) require the objective and constraint functions to be concave in the uncertain parameter as the reformulation maximizes over ζ . Also other adjustable robust optimization techniques in the literature as Fourier-Motzkin elimination, nonlinear decision rules, Benders decomposition, and the column-and-constraint generation (CCG) method of Zeng and Zhao (2013), are developed for linear adjustable problems and are not applicable for problems in which the adjustable variables occur in a nonlinear and convex way. We emphasize, however, that in this section we show that the dual of the nonlinear problem is linear in the adjustable variables, and hence for this dual problem the above mentioned well-known adjustable linear robust optimization techniques can be used.

Alternatively, finite adaptability approaches of Bertsimas and Dunning (2016) and Postek and den Hertog (2016) are effective for adjustable robust linear problems with integer wait-and-see decisions and polyhedral uncertainty sets, and can be readily extended to nonlinear cases. However, finite adaptability approaches often restrict adjustable variables to be piecewise constant, which can be too restrictive and may lead to trivial solutions, *e.g.*, it requires $A(x) = 0$ in (1) if \mathcal{U} is full dimensional. A possible remedy may be to eliminate the adjustable variables in equality constraints using Gaussian elimination, which is proven to be equivalent to imposing linear decision rules to the adjustable variables in the equality constraints (Zhen and den Hertog, 2017, Lemma 2).

In the next subsection, we propose a dual approach to derive an equivalent reformulation of (1). The reformulation is a two-stage robust linear optimization model, for which existing approaches for two-stage robust linear optimization can be used.

2.2. Consecutive dualization

To derive the linear reformulation of (1), we require (1) to satisfy the following property of *strong relatively complete recourse*.

ASSUMPTION 1 (Strong relatively complete recourse). *For all $x \in \mathcal{X}$ and all $\zeta \in \mathcal{U}$ there exists a $y \in \bigcap_{i=0}^{m_1} \text{ri}(\text{dom}(g_i))$, such that*

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

and for all $i = 1, \dots, m_1$ for which g_i is nonlinear we have $\zeta^\top F_i(x) + f_i(x) + g_i(y) < 0$. □

This assumption implies that each here-and-now decision is strictly feasible. This assumption is required to guarantee strong duality by Slater's condition in Theorem 1. It seems to be 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 linear optimization, see Birge and Louveaux (2011).

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:*

$$\inf_{x \in \mathcal{X}} \sup_{(u,v,w,z) \in \mathcal{V}} \inf_{\lambda \geq 0} \left\{ \sum_{i=0}^{m_1} v_i f_i(x) + d^\top \lambda - w^\top b(x) - \sum_{i=0}^{m_1} z_i \right. \\ \left. \sum_{k=1}^p D_{kj} \lambda_k \geq w^\top A_{\cdot j}(x) + \sum_{i=1}^{m_1} v_i F_{ij}(x), \quad j = 1, \dots, n_\zeta \right\}, \quad (5)$$

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

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

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:

$$\inf_{x \in \mathcal{X}} \sup_{\substack{\zeta \in \mathcal{U} \\ v \geq 0, w}} \inf_y f_0(x) + g_0(y) + \sum_{i=1}^{m_1} v_i (\zeta^\top F_i(x) + f_i(x) + g_i(y)) + w^\top (A(x)\zeta + By - b(x)). \quad (6)$$

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:

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \sup_{(u,v,w) \in \mathcal{W}} f_0(x) + \sum_{i=1}^{m_1} v_i (\zeta^\top F_i(x) + f_i(x)) + w^\top (A(x)\zeta - b(x)) - \sum_{i=0}^{m_1} v_i (g_i)^* \left(\frac{u_i}{v_i} \right), \quad (7)$$

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:

$$\inf_{x \in \mathcal{X}} \sup_{(u,v,w) \in \mathcal{W}} \inf_{\lambda \geq 0} \left\{ \sum_{i=0}^{m_1} v_i f_i(x) + d^\top \lambda - w^\top b(x) - \sum_{i=0}^{m_1} v_i (g_i)^* \left(\frac{u_i}{v_i} \right) \right. \\ \left. \sum_{k=1}^p D_{kj} \lambda_k \geq w^\top A_{\cdot j}(x) + \sum_{i=1}^{m_1} v_i F_{ij}(x), \quad j = 1, \dots, n_\zeta \right\}.$$

We then introduce epigraph variables z_i for every $v_i(g_i)^* \begin{pmatrix} u_i \\ v_i \end{pmatrix}$, $i = 0, \dots, m_1$, and finally obtain (5). \square

Note that the linear structure of the uncertainty set appears in the constraints of the dual formulation (5) and the convex structure of the adjustable variables is in the new uncertainty set of (5). In the specific linear case, where $F_i(x)$, $f_i(x)$ and $g_i(y)$ are affine functions, Theorem 1 coincides with the result in (Bertsimas and de Ruiter, 2016, Theorem 1). Also note that Theorem 1 holds even if the functions f_i and F_{ij} , $i = 0, \dots, m_1$, $j = 1, \dots, n_\zeta$, in model (1) are nonconvex in x .

In many cases \mathcal{V} is second-order cone representable (Lobo et al., 1998), making (5) 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-shelf 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). For several two-stage robust linear models, the structure of the optimal decision rules has been characterized. One can use Zhen et al. (2017b) to show that there exists a piecewise affine function that is optimal for λ in (5). More specifically, if \mathcal{U} is simplicial, there exists a linear decision rule that is optimal for λ in (5); if \mathcal{U} is a box, there exists a two-piecewise affine function that is optimal for λ in (5), and the techniques proposed in Gorissen and den Hertog (2013) and Ardestani-Jaafari and Delage (2016) can then be applied to solve Problem (5) approximately. Even when the structure of optimal decision rules is known, it is often difficult to find optimal solutions due to the computational intractability of such rules.

The main benefit (and purpose) of the dual formulation is that the objective and constraint functions are now linear in the adjustable variables and uncertain parameters, which can therefore be solved with any existing method applicable to two-stage linear robust models. For instance, beside the approximative linear decision rules, exact methods for two-stage robust linear problems with polyhedral uncertainty sets such as column/constraint generation methods of Zeng and Zhao (2013), Ayoub and Poss (2016), Simchi-Levi et al. (2018) and Benders' decomposition method of Jiang et al. (2013) can be readily extended to solve the new dual formulation, which then requires of solving mixed integer nonlinear optimization problems in each iteration. However, existing solvers may have many difficulties in solving mixed integer convex nonlinear problems efficiently. Furthermore, finite adaptability approaches of Bertsimas and Dunning (2016) and Postek and den Hertog (2016) can be extended to solve (5). These approaches are effective for adjustable robust linear problems with integer wait-and-see decisions and polyhedral uncertainty sets, while for general convex uncertainty sets this type of approach may take infinite number of iterations to obtain the optimal solution. Approaches based on copositive reformulations are recently proposed by Xu and Burer (2018), Hanasusanto and Kuhn (2018) and Hanasusanto and Xu (2018). Unfortunately, it is

not clear how to extend their approach to deal with general convex functions in problems (1) and (5). In the remainder of this paper, we focus on one of the most prominent methods, that is, linear decision rules (Ben-Tal et al., 2004).

2.3. Linear decision rule solution for the dual problem

Firstly, substituting the wait-and-see variables in (5) with the linear decision rule:

$$\lambda(u, v, w, z) = \sum_{i=0}^{m_1} \Psi_i^\top u^i + \sum_{i=0}^{m_1} t_i v_i + \Phi^\top w + \sum_{i=0}^{m_1} \eta_i z_i,$$

where $\Psi_i = [\psi_i^1 \dots \psi_i^k] \in \mathbb{R}^{n_y \times k}$, $t_i, \eta_i \in \mathbb{R}^k$, and $\Phi = [\phi_i^1 \dots \phi_i^k] \in \mathbb{R}^{m_2 \times k}$ for all $i = 0, \dots, m_1$, yields the following static robust optimization problem:

$$\begin{aligned} & \inf_{\substack{x \in \mathcal{X}, t_i \\ \Psi_i, \Phi, \eta_i}} \sup_{(u, v, w, z) \in \mathcal{V}} v^\top F_0(x) + d^\top (\lambda_0 + \Lambda u + \Gamma v + \Omega w + \Pi z) + w^\top (A^0 x - b^0) - e^\top z_i \\ & \text{s.t. } \forall (u, v, w, z) \in \mathcal{V}: \\ & \left\{ \begin{aligned} & D_{\cdot j}^\top \left(\sum_{i=0}^{m_1} \Psi_i^\top u^i + \sum_{i=0}^{m_1} t_i v_i + \Phi^\top w + \sum_{i=0}^{m_1} \eta_i z_i \right) \geq w^\top (A^j x - b^j) + v^\top [F_{\cdot j}^0(x)] \quad j = 1, \dots, n_\zeta \\ & \sum_{i=0}^{m_1} \Psi_i^\top u^i + \sum_{i=0}^{m_1} t_i v_i + \Phi^\top w + \sum_{i=0}^{m_1} \eta_i z_i \geq 0, \end{aligned} \right. \quad (8) \end{aligned}$$

where $F_0(x) = (f_0(x), \dots, f_{m_1}(x))^\top \in \mathbb{R}^{m_1+1}$, $F_{\cdot j}(x) \in \mathbb{R}^{m_1}$ and $D_{\cdot j} \in \mathbb{R}^p$ are the j -th column vectors of $F(x)$ and D , respectively. Since the uncertain parameters appear linearly in the static robust optimization problem (8), and \mathcal{V} is convex, one can use standard robust optimization techniques to derive the following tractable reformulation:

$$\inf_{\substack{x \in \mathcal{X}, y_0, y_j \\ t, \gamma, \eta \geq 0, \tau, \Psi_i}} \left\{ \begin{aligned} & f_0(x) + \tau \\ & \gamma_{00} g_0 \left(\frac{y'_{i0}}{\gamma_{00}} \right) + \sum_{k=1}^p d_k t_{0k} \leq \tau \\ & \gamma_{0j} g_0 \left(\frac{y'_{ij}}{\gamma_{0j}} \right) \leq \sum_{k=1}^p D_{kj} t_{0k}, \quad j = 1, \dots, n_\zeta \\ & f_i(x) + \gamma_{i0} g_i \left(\frac{y'_{i0}}{\gamma_{i0}} \right) + \sum_{k=1}^p d_k t_{ik} \leq 0, \quad i = 1, \dots, m_1 \\ & F_{ij}(x) + \gamma_{ij} g_i \left(\frac{y'_{ij}}{\gamma_{ij}} \right) \leq \sum_{k=1}^p D_{kj} t_{ik}, \quad i = 1, \dots, m_1 \\ & j = 1, \dots, n_\zeta \\ & B y_0 + \sum_{k=1}^p d_k \psi_0^k = b(x) \\ & A_{\cdot j}(x) + B y_j = \sum_{k=1}^p D_{kj} \psi_0^k, \quad j = 1, \dots, n_\zeta \\ & \gamma_{i0} + \sum_{k=1}^p \eta_{ik} d_k = 1, \quad i = 0, \dots, m_1 \\ & \gamma_{ij} = \sum_{k=1}^p D_{kj} \eta_{ik}, \quad i = 0, \dots, m_1, \quad j = 1, \dots, n_\zeta \\ & y'_{i0} = y_0 - \sum_{k=1}^p d_k \psi_i^k, \quad i = 0, \dots, m_1 \\ & y'_{ij} = y_j + \sum_{k=1}^p D_{kj} \psi_i^k, \quad i = 0, \dots, m_1, \quad j = 1, \dots, n_\zeta \end{aligned} \right. \quad (9)$$

Maybe interestingly from a practitioner's perspective, we find that this tractable reformulation does not contain the conjugates g_i^* , but only the original functions g_i for every $i = 0, \dots, m_1$. Hence, there is no need to derive conjugate functions.

2.4. Primal decision rule solution obtained from primal solution

Consider the following primal linear decision rule

$$y(\zeta) = y_0 + \sum_{j=1}^{n_\zeta} y_j \zeta_j, \quad (10)$$

where $y_0, \dots, y_{n_\zeta} \in \mathbb{R}^{n_y}$. A conservative approximation of the primal formulation (1) using this linear decision rule is the following:

$$\inf_{\substack{x \in \mathcal{X}, \tau \\ y_0, y_j, \Psi_i}} \sup_{\zeta \in \mathcal{U}} \left\{ f_0(x) + \tau \left| \begin{array}{l} g_0(y(\zeta)) \leq \tau \\ \zeta^\top F_i(x) + f_i(x) + g_i(y(\zeta)) \leq 0, \quad i = 1, \dots, m_1 \\ A(x)\zeta + By(\zeta) = b(x) \end{array} \right. \right\}. \quad (11)$$

Note that the problem (11) itself is generally computationally intractable, because the inner maximization problem tries to maximize a convex function over a polyhedron, which is in general NP-hard. In this section we show how to obtain a primal solution to (11) from the dual solution of (9). To be precise, we show in this section that linear decision rule (10) is feasible for (11) if the values y_j , $j = 1, \dots, n_\zeta$, are taken from the optimal solution of (9). To derive this result, we first introduce the following result, which we term the ‘‘perspective relaxation’’.

LEMMA 1. *If the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, we have:*

$$f \left(\sum_{i=1}^N \alpha_i x_i \right) \leq \sum_{i=1}^N \alpha_i \gamma_i f \left(\frac{x_i}{\gamma_i} \right), \quad (12)$$

where $\sum_{i=1}^N \alpha_i \gamma_i = 1$, and $\alpha_i, \gamma_i \geq 0$ for all $i = 1, \dots, N$.

For ease of derivation, we rewrite the linear decision rule (10) as follows:

$$\begin{aligned} y(\zeta) &= y_0 + \sum_{j=1}^{n_\zeta} y_j \zeta_j - \underbrace{\sum_{k=1}^p \psi_i^k (d - D\zeta)_k + \sum_{k=1}^p \psi_i^k (d - D\zeta)_k}_{=0} \\ &= y_0 - \underbrace{\sum_{k=1}^p d_k \psi_i^k}_{y'_{i0}} + \sum_{j=1}^{n_\zeta} \underbrace{\left(y_j + \sum_{k=1}^p D_{kj} \psi_i^k \right)}_{y'_{ij}} \zeta_j + \sum_{k=1}^p \psi_i^k (d - D\zeta)_k \quad \forall i = 1, \dots, n_\zeta \quad (13) \\ &= y_i(\zeta) \quad \forall i = 1, \dots, n_\zeta, \end{aligned}$$

where ψ_i^k for all $i = 0, \dots, n_\zeta$ and $k = 1, \dots, p$ are taken from the optimal solution of (9). Although it appears that the right-hand-side depends on i , one can see that all terms depending on i cancel. Hence, (13) and (10) give the same values for the wait-and-see decision $y(\zeta)$ for all $\zeta \in \mathcal{U}$. Furthermore, take γ_{ij} and η_{ij} for $i = 0, \dots, n_\zeta$ and $j = 1, \dots, n_\zeta$ also from the optimal solution of (9) so

that $\gamma_{i0} + \sum_{k=1}^p \eta_{ik} d_k = 1$. Using the first constraint in (11), we can derive the following result for all $\zeta \in \mathcal{U}$

$$\begin{aligned} g_0(y_0(\zeta)) &= g_0 \left(y'_{00} + \sum_{j=1}^{n_\zeta} y'_{0j} \zeta_j + \sum_{k=1}^p \psi_0^k (d - D\zeta)_k \right) \\ &\leq \gamma_{00} g_0 \left(\frac{y'_{00}}{\gamma_{00}} \right) + \sum_{j=1}^{n_\zeta} \gamma_{0j} \zeta_j g_0 \left(\frac{y'_{0j}}{\gamma_{0j}} \right) + \sum_{k=1}^p \eta_{0k} (d - D\zeta)_k g_0 \left(\frac{\psi_0^k}{\eta_{0k}} \right), \end{aligned}$$

where for the last inequality we have used Lemma 1. Hence, if the inequality

$$\gamma_{00} g_0 \left(\frac{y'_{00}}{\gamma_{00}} \right) + \sum_{j=1}^{n_\zeta} \gamma_{0j} \zeta_j g_0 \left(\frac{y'_{0j}}{\gamma_{0j}} \right) + \sum_{k=1}^p \eta_{0k} (d - D\zeta)_k g_0 \left(\frac{\psi_0^k}{\eta_{0k}} \right) \leq \tau$$

holds for all $\zeta \in \mathcal{U}$, then so does $g_0(y_0(\zeta)) \leq \tau$. We can apply the perspective relaxation for the other constraints in (11) analogous to the first constraint as done above. The result is that, if y'_{ij} , ψ_i^k , γ_{ij} , η_{ij} for all $k = 1, \dots, p$, $i = 0, \dots, n_\zeta$ and $j = 1, \dots, n_\zeta$ satisfy the following set of inequalities

$$\forall \zeta \in \mathcal{U} \begin{cases} \gamma_{00} g_0 \left(\frac{y'_{00}}{\gamma_{00}} \right) + \sum_{j=1}^{n_\zeta} \gamma_{0j} \zeta_j g_0 \left(\frac{y'_{0j}}{\gamma_{0j}} \right) + \sum_{k=1}^p \eta_{0k} (d - D\zeta)_k g_0 \left(\frac{\psi_0^k}{\eta_{0k}} \right) \leq \tau \\ \zeta^\top F_i(x) + f_i(x) + \gamma_{i0} g_i \left(\frac{y'_{i0}}{\gamma_{i0}} \right) + \dots \\ + \sum_{j=1}^{n_\zeta} \gamma_{ij} \zeta_j g_i \left(\frac{y'_{ij}}{\gamma_{ij}} \right) + \sum_{k=1}^p \eta_{ik} (d - D\zeta)_k g_i \left(\frac{\psi_i^k}{\eta_{ik}} \right) \leq 0, \quad i = 1, \dots, m_1 \\ \gamma_{i0} + \sum_{j=1}^{n_\zeta} \gamma_{ij} \zeta_j + \sum_{k=1}^p \eta_{ik} (d - D\zeta)_k = 1, \quad i = 0, \dots, m_1 \\ A(x)\zeta + B \left(y_0 + \sum_{j=1}^{n_\zeta} y_j \zeta_j \right) = b(x), \quad \gamma_{i0} + \sum_{k=1}^p \eta_{ik} d_k = 1, \quad i = 0, \dots, m_1 \\ \gamma_{ij} = \sum_{k=1}^p D_{kj} \eta_{ik}, \quad i = 0, \dots, m_1, \quad j = 1, \dots, n_\zeta \\ y'_{i0} = y_0 - \sum_{k=1}^p d_k \psi_i^k, \quad i = 0, \dots, m_1 \\ y'_{ij} = y_j + \sum_{k=1}^p D_{kj} \psi_i^k, \quad i = 0, \dots, m_1, \quad j = 1, \dots, n_\zeta, \end{cases} \quad (14)$$

then they also satisfy the constraints from (11) for all $\zeta \in \mathcal{U}$. The set of inequalities (14) form a set of robust constraints, all of which are linear in the uncertain parameter. Therefore, using standard techniques for robust optimization one can deduce its robust counterpart that is equivalent to the set of constraints given in (9). Hence, once an optimal (or feasible) solution for (11) is found, the optimal values for $y_0, y_1, \dots, y_{n_\zeta}$ form a primal linear decision rule (10) that is feasible for the primal problem (11).

In a nutshell, we have generalized the results of Bertsimas and de Ruiter (2016) to nonlinear cases, i.e., we have shown that the primal linear decision rules can be obtained from the dual solution. The only remaining open question is whether an optimal solution for (9) also yields the optimal primal linear decision rule. Optimality can be established for linear cases, as is done in Bertsimas and de Ruiter (2016), but this does not extend to the nonlinear case. In fact, the numerical results in Section §4 show that the optimal primal linear decision rule, obtained via extensive computer power, might be slightly better than the primal linear decision rule that one recovers from (9). Hence, the equivalence between primal and dual linear decision rules proven

in Bertsimas and de Ruiter (2016) cannot be extended to the nonlinear case. Nevertheless, the numerical results empirically show that our linear decision rules for nonlinear cases are on par with the good performance seen in fully linear adjustable robust problems.

2.5. Extension to multistage robust nonlinear models

To accomplish multistage robust solutions, the wait-and-see decision must comply with non-anticipativity. That is, a decision in some period l can only rely on the information that is available up to that period. For instance, the values for the uncertain parameter known at time l might be ζ_1, \dots, ζ_l , whereas the values of $\zeta_{l+1}, \dots, \zeta_{n_\zeta}$ are still uncertain. Therefore, the nonanticipative version the linear decision rule (10) is given by

$$y(\zeta) = y_0 + \sum_{j=1}^{n_\zeta} y_j \zeta_j$$

$$(y_j)_l = 0 \quad l = 1, \dots, j-1, \quad j = 2, \dots, n,$$

so the l -th entry of the vector y_j is zero for the values of l indicated above. To enforce a nonanticipative linear decision in the primal we therefore only have to include these constraints in (9). The same reasoning naturally holds for problems where the original primal model is linear, but this is however not noted in Bertsimas and de Ruiter (2016).

2.6. Extension to non-polyhedral uncertainty sets

The format of the primal models is only given for polyhedral uncertainty sets. However, for ellipsoidal (or any second-order cone representable) uncertainty sets polyhedral outer approximations from Ben-Tal and Nemirovski (2001) could be used to enable the use of the approach proposed in the current paper. An attractive property of these outer approximations is that the number of linear constraints to approximate an ellipsoid is polynomial in the problem dimensions and $|\ln(\epsilon)|$, where ϵ is the accuracy of the approximation. Problem (5) then no longer yields an equivalent problem to the original problem, but a safe approximation. In practice, however, the choice of scenarios for which one needs protection is often to some extent unknown, which means that one does not need a polyhedral approximation with high accuracy.

We also propose an other way to extend our approach to non-polyhedral uncertainty sets. Zhen et al. (2018) extends the reformulation linearization technique for bilinear formulations to cases where the feasible region is non-polyhedral. Hence, in case \mathcal{U} is non-polyhedral, we could apply this extended reformulation linearization technique to (7). In Zhen and de Ruiter (2019), the authors further extend the reformulation linearization technique to solve multistage problems. The resulting optimization problem is computationally tractable given that \mathcal{U} is “tractable”. For example, if \mathcal{U}

is ellipsoidal, then the resulting optimization problem contains a conic quadratic part because of the ellipsoidal uncertainty set. However, extensive numerical experiments should be carried out to analyze the performance of this extended reformulation linearization technique for problems with non-polyhedral uncertainty sets.

3. Bounds on the optimal value

The dualized model (5) 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 $\{\zeta^1, \dots, \zeta^S\}$ 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^s to each scenario ζ^s , for $s = 1, \dots, S$. The lower bound model is therefore the “sampled version” of the original model:

$$\begin{aligned}
 & \inf_{\substack{\tau, x \in \mathcal{X} \\ y^1, \dots, y^S}} \tau \\
 & \text{s.t.} \quad f_0(x) + g_0(y^s) \leq \tau \quad \forall s = 1, \dots, S \\
 & \quad (\zeta^s)^\top F_i(x) + f_i(x) + g_i(y^s) \leq 0 \quad \forall i = 1, \dots, m_1, s = 1, \dots, S \\
 & \quad A(x)\zeta^s + By^s = b(x) \quad \forall s = 1, \dots, S.
 \end{aligned} \tag{15}$$

Model (15) 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 (5). In the dualized model we would take a finite subset $\{(u^1, w^1, v^1, z^1), \dots, (u^S, w^S, v^S, z^S)\}$ from \mathcal{V} with a

single optimization variable λ^s for each scenario (u^s, w^s, v^s, z^s) , $s = 1, \dots, S$. The sampled version of the dualized model is:

$$\begin{aligned} & \inf_{\substack{\tau, x \in \mathcal{X} \\ \lambda^1, \dots, \lambda^S \geq 0}} \tau \\ \text{s.t.} \quad & f_0(x) + \sum_{i=1}^{m_1} v_i^s f_i(x) + d^\top \lambda^s - (w^s)^\top b(x) - \sum_{i=0}^{m_1} z_i^s \leq \tau \quad \forall s = 1, \dots, S \\ & \sum_{s=1}^p D_{kj} \lambda_k^s \geq (w^s)^\top A_{\cdot j}(x) + \sum_{i=1}^{m_1} (v_i^s) F_{i,l}(x) \quad \forall j = 1, \dots, n_\zeta, \quad s = 1, \dots, S, \end{aligned} \quad (16)$$

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 2. *Let \mathcal{U} be a polyhedral uncertainty set with S extreme points ζ^1, \dots, ζ^S . Then the optimal here-and-now solution \bar{x} of model (15) is also optimal for model (1) and their optimal objective values coincide.*

Proof. Let $\bar{\tau}, \bar{x}, \bar{y}^1, \dots, \bar{y}^S$ be the optimal solution of (15). We know that the optimal value $\bar{\tau}$ of the sampled model (15) 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} :

$$\zeta = \sum_{s=1}^S \alpha_s \zeta^s \quad (17)$$

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

$$y = \sum_{s=1}^S \alpha_s \bar{y}^s, \quad (18)$$

with $\alpha_1, \dots, \alpha_S$ the same values as those in the convex combination of (17). Then we have:

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

where the first inequality is due to convexity of the functions g_i and the last inequality is due to the fact that $\bar{x}, \bar{y}^1, \dots, \bar{y}^S$ is feasible for (15). Analogously, we can show that for \bar{x} and decision rule y from (18) 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 $\bar{\tau}$. \square

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 $\{(u^1, w^1, v^1, z^1), \dots, (u^S, w^S, v^S, z^S)\}$ 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 $\{\zeta^1, \dots, \zeta^S\}$ from the original model and the dual scenarios $\{(u^1, w^1, v^1, z^1), \dots, (u^S, w^S, v^S, z^S)\}$ by using a dual approach. By dualizing over $\lambda_1, \dots, \lambda_K$ we get the following equivalent formulation of (16):

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \sup_{1 \leq s \leq S} f_0(x) + \sum_{i=1}^{m_1} v_i^s (\zeta^\top F_i(x) + f_i(x)) + (A(\zeta)x - b(\zeta))^\top w^s - \sum_{i=0}^{m_1} z_i^s, \quad (19)$$

which is similar to (7), but with the inner supremum over $(u, w, v, z) \in \mathcal{V}$ replaced by the finite subset with S scenarios. For a fixed x we can obtain primal scenarios ζ^s for each s as the maximizers of model (19):

$$\zeta^s \in \arg \max_{\zeta \in \mathcal{U}} \left\{ \sum_{i=1}^{m_1} v_i^s (\zeta^\top F_i(x)) + (w^s)^\top (A(\zeta)x - b(\zeta)) \right\}. \quad (20)$$

The resulting set of scenarios $\{\zeta^1, \dots, \zeta^S\}$ can then be used in the sampled model (15). One can now solve either the primal sampled model (15), which is a convex optimization model, or the dual sampled model (16), 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 (15) or (16) 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 (20), then we can show that the lower bound from (15) is always higher (or equal to) (16). 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 $F_i = 0$ for all $i = 1, \dots, m_1$ and there exists $\bar{A} \in \mathbb{R}^{m_2 \times n_\zeta}$ such that for all x we have $A(x) = \bar{A}$.

Two-stage robust linear problems with right-hand-side uncertainties are investigated in recent papers, *e.g.*, Xu and Burer (2018) and Hanasusanto and Xu (2018). Note that in the case of right-hand-side uncertainty, the scenarios ζ^s can be obtained in (20) independent of the here-and-now decision x as the only terms depending on ζ are $(w^s)^\top b(\zeta)$. In the following theorem, we exploit this observation to compute a set of dominating scenarios for the primal nonlinear models, and tighten the obtained lower bounds from the dual scenarios.

THEOREM 3. *Let $\{(u^1, w^1, v^1, z^1), \dots, (u^S, w^S, v^S, z^S)\}$ be a finite set of dual scenarios and $\{\zeta^1, \dots, \zeta^S\}$ be a set of primal scenarios obtained from (20). If there is only right-hand-side uncertainty in model (1), then the lower bound from (15) is at least as tight as the lower bound from (16).*

Proof. By duality for linear programming, (16) is equivalent to (19). The latter formulation can be written as

$$\inf_{x \in \mathcal{X}} \sup_{s \in \{1, \dots, S\}} \left\{ f_0(x) + \sum_{i=1}^{m_1} v_i^s f_i(x) + (w^s)^\top (\bar{A}x - b(\zeta^s)) - \sum_{i=0}^{m_1} z_i^s \right\}, \quad (21)$$

where ζ^s are the primal scenarios obtained by (20). Since (u^s, w^s, v^s, z^s) are in \mathcal{V} for all $s = 1, \dots, S$, the value of (21) must be smaller than or equal to

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

since we maximize over (u^s, w^s, v^s, z^s) in \mathcal{V} , instead of a fixing these S values beforehand. The value of this optimization problem is, by dualizing over (u^s, w^s, v^s, z^s) , equivalent to (15). Hence, the optimal objective value of the model (15) is at least as high as the optimal objective value of (16). \square

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.*, (4) with $F_i = 0$ for all $i = 1, \dots, m_1$, $A^i(x) = \bar{A}^i$ for all $i = 0, \dots, m_1$, and $A(x) = \bar{A}$. Theorem 3 only holds when we have right-hand-side uncertainty in Problem (1). In Appendix B, we show via an example that the lower bound from the sampled version of the dual model (16) can be higher than

the lower bound resulting from the sampled version of the primal problem (15) when there is left-hand-side uncertainty. In Appendix C, we evaluate the performance of the lower bounding scheme proposed in this subsection with the technique of Bertsimas and de Ruiter (2016) using the same numerical experiment considered in Bertsimas and de Ruiter (2016). The lower bounds obtained from our method are much tighter than the ones obtained using the technique of Bertsimas and de Ruiter (2016).

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 , $i, j = 1, \dots, N$, 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 , $i = 1, \dots, N$, 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, $i, j = 1, \dots, N$. 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} , $i, j = 1, \dots, N$. This distribution model can now be written as a specific instance of the primal problem as follows:

$$\inf_{\substack{x \in \mathcal{X} \\ z, \tau}} \sup_{\zeta \in \mathcal{U}} \inf_{y \geq 0} \left\{ \sum_{i=1}^N c_i x_i + \tau \left| \begin{array}{l} \sum_{i,j=1}^N t_{ij} y_{ij} + \frac{1}{2} \sum_{i,j=1}^N t_{ij} (y_{ij} - z_{ij})^2 \leq \tau \\ \sum_{j=1}^N y_{ji} - \sum_{j=1}^N y_{ij} \geq \zeta_i - x_i \quad i = 1, \dots, N \end{array} \right. \right\}, \quad (22)$$

where the third term in the objective of (22) captures the cost of contract violation, and $\mathcal{X} = \{x \in \mathbb{R}_+^N \mid e^\top x \geq \Gamma, x_i \leq K_i \quad i = 1, \dots, N\}$. The constraints in (22) 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 , $i = 1, \dots, N$, 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:

$$\inf_{\substack{x \in \mathcal{X} \\ z, \tau}} \sup_{(u,v,w) \in \mathcal{V}} \inf_{\lambda \geq 0} \left\{ \sum_{i=1}^N c_i x_i + \tau \left| \begin{array}{l} \sum_{i=1}^N (\hat{\zeta}_i \lambda_i - u_i x_i) + \Gamma \lambda_0 \dots \\ \quad - \sum_{i,j=1}^N [(u_j - u_i - t_{ij} - v_{ij}) z_{ij} + \frac{1}{2} w_{ij}] \leq \tau \\ \lambda_0 + \lambda_i \geq u_i \quad i = 1, \dots, N \end{array} \right. \right\}, \quad (23)$$

where $\mathcal{V} = \{(u, v, w) \geq 0 : (u_i - u_j + v_{ij} - t_{ij})^2 \leq w_{ij}t_{ij} \quad \forall i, j = 1, \dots, N\}$. Note that in both problem formulations (22) and (23), the epigraphical auxiliary variable τ can be eliminated, then it can be verified that the resulting formulations satisfy (strongly relative) complete recourse. Now we apply linear decision rules, Fourier-Motzkin elimination and column-and-constraint generation method of Zeng and Zhao (2013) to solve (23).

4.2. Numerical setting

We choose $N \in \{5, 10, 20, 30, 40, 50, 60\}$ 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 $i = 1, \dots, \lceil N/10 \rceil + 1$ warehouses and 10 for $i = \lceil N/10 \rceil + 1, \dots, N$ stores. The individual maximum demand $\hat{\zeta}$ and the capacity K_i , $i = 1, \dots, N$, 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 (23) 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>). All the reported numbers in the tables are the average of 10 randomly generated instances.

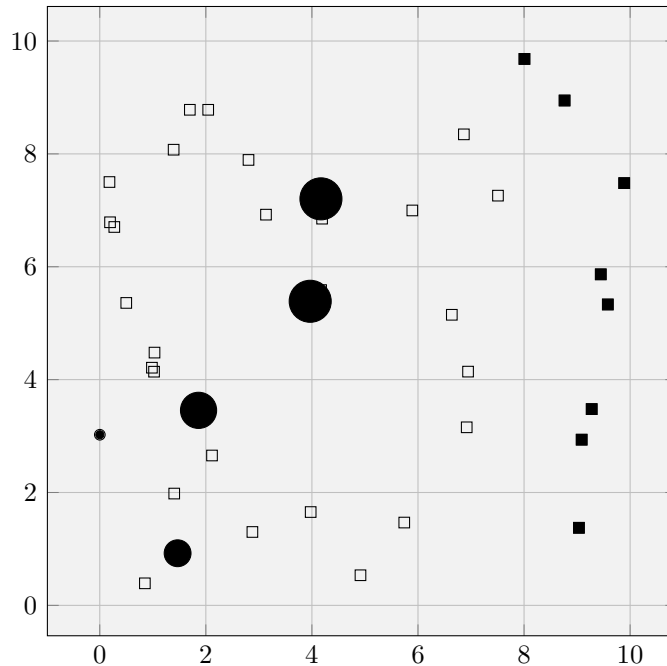


Figure 1 Stock allocation for $N = 40$ with 35 stores (squares) and 5 warehouses (circles) for one random instance. The filled dots have stock and the larger the dots are, the more stock is allocated.

4.3. Results on Fourier-Motzkin elimination

We first consider a small instance and present the results in Table 1. We follow the suggestion of Zhen et al. (2017b): we first eliminate the adjustable variables that produce the least number of constraints. Hence, we try to eliminate as many adjustable variables as possible while keeping the problem at its minimal size. One can observe that the solutions converge to optimality as more adjustable variables in (23) 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 (22) because the adjustable variables appear nonlinearly in the model. By solving (22) and (23) 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 (22) and (23) 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 Problem (22) 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). Hence, we only focus on the LB-P lower bounds for the rest of this paper.

We also numerically evaluate the effect of “perspective relaxation” in Lemma 1. By solving the corresponding reformulation (11) of (22) via vertex enumeration, we obtain an average optimal value 665, which constitutes an upper bound of the true average optimal value 635 (see Table 1). The upper bound from solving the corresponding reformulation (11) of (22) (*i.e.*, before “perspective relaxation”) is indeed much tighter than solving (22) with linear decision rules (*i.e.*, after “perspective relaxation”).

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 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, where P-S and D-L are obtained from solving (22) with static decision rules and (23) with linear decision rules, respectively.

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.

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 (22) and (23) with static decision rules, respectively. D-L is obtained from solving (23) 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.

#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	703	703	696	682	666	654	635
Time(s)	0.03	0.03	0.03	0.04	0.07	0.15	0.66
LB-P	632	633	634	635	635	635	635
LB-D	49	212	385	502	609	622	635

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 (22) with static decision rules. D-L is obtained from solving (23) 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.

#Elim.	0	1	2	3	4	5	6	7	8	9	10	11
#Constr.	22	21	21	23	29	43	73	135	261	515	1025	149424
N=10 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	*
#Constr.	42	41	41	43	49	63	93	165	281	535	1045	2067
N=20 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 (22) with static decision rules. D-L is obtained from solving (23) 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	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

4.4. Results on column-and-constraint generation method of Zeng and Zhao (2013)

Note that the primal problem (22) cannot be solved using the column-and-constraint generation (CCG) method of Zeng and Zhao (2013) because the resulting subproblem constitutes a convex

maximization problem with binary variables. However, by applying Theorem 1 to (22), the obtained dual problem (23) can be directly solved using CCG method of Zeng and Zhao (2013), and the corresponding subproblem is a convex minimization problem with binary variables, which can be solved using MOSEK ApS (2017).

Due to the excessive computational effort of solving mixed integer nonlinear optimization problems, we consider relatively small instances and report the results in Table 4. The obtained lower bounds are again remarkably tight, *i.e.*, the average optimality gap is less than 0.5%. OPT denotes the optimal objective value obtained from the exact method of Zeng and Zhao (2013) using the lower bound solutions as warm start. We refer to Zhen and de Ruiter (2019) for a detailed evaluation of the effect of using such a warm start for the exact method of Zeng and Zhao (2013).

Table 4 Lot-sizing problem with $N \in \{5, 7, 9, 11, 13, 15\}$. P-S is obtained from solving (22) with static decision rules. D-L is obtained from solving (23) with linear decision rules. LB-P denotes the lower bounds obtained from the primal scenarios (see §3.3). OPT is obtained from solving (23) via Zeng and Zhao (2013). Time(s) reports the computation time (in seconds) for solving OPT.

N	5	7	9	11	13	15
P-S	822	1211	1604	2000	2398	2690
D-L	677	837	967	1083	1171	1211
LB-P	605	744	887	981	1064	1113
OPT	607	746	890	984	1069	1119
Time(s)	1	4	18	289	1492	3370

5. Example 2: worst-case energy configuration 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 , respectively. The tension in spring i is a nonlinear function of the distance between its endpoints, *i.e.*, $\|x_i - x_{i+1}\|_2$:

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

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.*,

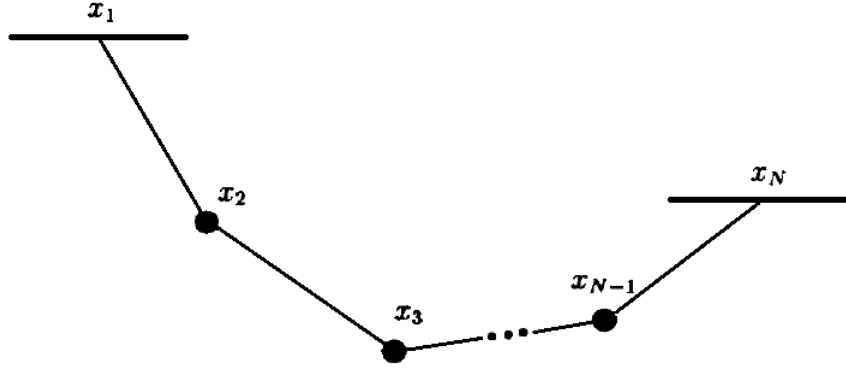


Figure 2 System of nodes (weights) connected by springs from Lobo et al. (1998). The first and last node positions, *i.e.*, x_1 and x_N , are fixed.

values of x_1, \dots, 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:

$$\begin{aligned} \inf_{x \geq 0} \quad & w \sum_{i=1}^N x_{i2} + \frac{s}{2} \sum_{i=1}^{N-1} \left[(\|x_i - x_{i+1}\|_2 - l_i^0)_+ \right]^2 \\ \text{s.t.} \quad & x_1 = a, \quad x_N = b, \end{aligned} \quad (24)$$

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:

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

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

$$\inf_{x \geq 0} \sup_{\zeta \in \mathcal{U}} \left\{ w \sum_{i=1}^N x_{i2} + \frac{s}{2} \sum_{i=1}^{N-1} \left[(\|x_i - x_{i+1}\|_2 - l_i(\zeta))_+ \right]^2 \mid x_1 = a, \quad x_N = b \right\}, \quad (25)$$

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

$$\inf_{x \geq 0} \sup_{\zeta \in \mathcal{U}} \inf_{y \geq 0} \left\{ w \sum_{i=1}^N x_{i2} + \frac{s}{2} \sum_{i=1}^{N-1} y_i^2 \mid \begin{array}{l} \|x_i - x_{i+1}\|_2 - l_i(\zeta) \leq y_i \quad i = 1, \dots, N-1 \\ x_1 = a, \quad x_N = b \end{array} \right\}. \quad (26)$$

It can be verified that models (25) and (26) are equivalent, that is, eliminating all the y_i 's for $i = 1, \dots, N-1$ in (26) we obtain (25). We solve the dualized formulation of (26) 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 $U(0, 4)$, $i = 1, \dots, N - 1$, 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 $i = 1, \dots, N - 1$, 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 (26) 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 5 shows that the approximations from solving the dualized model of (26) 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 (26) 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 5 Equilibrium of system with $N - 1$ piecewise-linear springs for $N \in \{15, 20, 30, 45, 65, 100\}$. P-S is obtained from solving (26) with static decision rules. D-L is obtained from solving the dualized model of (26) 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

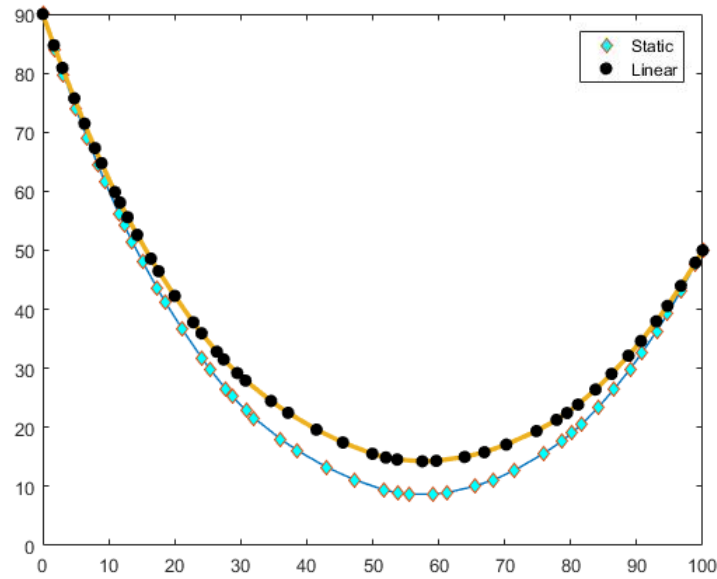


Figure 3 System of nodes (weights) connected by 44 springs for $N = 45$. The diamonds and dots represent the robust locations of the nodes from solving (26) and its dualized formulation with static decision rules and linear decision rules, respectively.

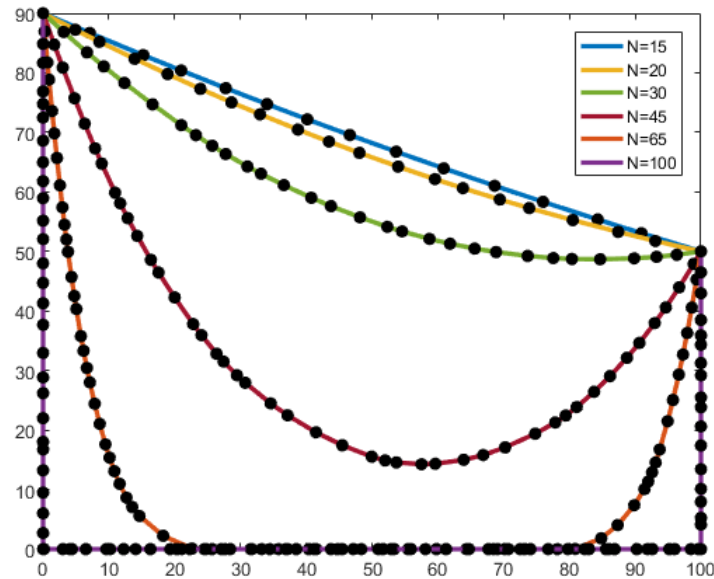


Figure 4 System of nodes (weights) connected by $N - 1$ springs for $N \in \{15, 20, 30, 45, 65, 100\}$. The dots represent the robust location of the nodes.

6. Conclusions 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. As a final note we

want to emphasize that our method uses duality for the recourse decisions, so it is not straightforward to extend our method to problems with integer recourse. Also, models need to be formulated in such a way that they have strong relatively complete recourse to allow strong duality to be guaranteed. 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).

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 this new approach. Besides the exact method of Zeng and Zhao (2013), the methods developed in Bertsimas and Dunning (2016), Postek and den Hertog (2016), and Simchi-Levi et al. (2018) may also be combined with the proposed approach to further improve the obtained solutions.

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 partially supported by the NWO Grant 613.001.208.

A. Proof for Lemma 1

Proof. Since $\sum_{i=1}^N \alpha_i \gamma_i = 1$, we first rewrite $f(\sum_{i=1}^N \alpha_i x_i)$ into its perspective form and denote it as $g: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$:

$$\begin{aligned}
\left(\sum_{i=1}^N \alpha_i \gamma_i \right) f \left(\frac{\sum_{i=1}^N \alpha_i x_i}{\sum_{i=1}^N \alpha_i \gamma_i} \right) &= g \left(\sum_{i=1}^N \begin{pmatrix} \alpha_i x_i \\ \alpha_i \gamma_i \end{pmatrix} \right) \\
&= g \left(\sum_{i=1}^N \alpha_i \gamma_i \begin{pmatrix} x_i / \gamma_i \\ 1 \end{pmatrix} \right) \\
&\leq \sum_{i=1}^N \alpha_i \gamma_i g \left(\begin{pmatrix} x_i / \gamma_i \\ 1 \end{pmatrix} \right) \\
&= \sum_{\substack{i=1: \\ \gamma_i > 0}}^N \alpha_i \gamma_i f \left(\frac{x_i}{\gamma_i} \right) + \sum_{\substack{i=1: \\ \gamma_i = 0}}^N \alpha_i \delta_{\text{dom } f^*}^*(x_i) \\
&= \sum_{\substack{i=1: \\ \gamma_i > 0}}^N \alpha_i \gamma_i f \left(\frac{x_i}{\gamma_i} \right),
\end{aligned}$$

where the inequality follows from the convexity of g and $\sum_{i=1}^N \alpha_i \gamma_i = 1$ for $\alpha_i, \gamma_i \geq 0$, $i = 1, \dots, N$, while the third equality is due to the definition of perspective function. \square

B. Example when conditions in Theorem 3 are violated

Consider the following problem:

$$\min_{(x_1, x_2) \in \mathcal{X}} \max_{(\zeta_1, \zeta_2) \in \mathcal{U}} \min_y \{ -y \mid -1 + x_1 \zeta_1 + x_2 \zeta_2 + y^2 \leq 0 \},$$

where $\mathcal{X} = \{(x_1, x_2) \geq 0 \mid x_1 + x_2 = 1\}$ and $\mathcal{U} = \{(\zeta_1, \zeta_2) \geq 0 \mid \zeta_1 + \zeta_2 \leq 1\}$. This problem satisfies the strong relatively complete recourse condition, because for all $(x_1, x_2) \in \mathcal{X}$ and $(\zeta_1, \zeta_2) \in \mathcal{U}$ the wait-and-see decision $y = 0$ is feasible. Using Theorem 1 we can obtain the dual version of this problem:

$$\min_{(x_1, x_2) \in \mathcal{X}} \max_{v \geq 0} \min_{\lambda \geq 0} \left\{ -\frac{1}{4v} - v + \lambda \mid \lambda \geq vx_1, \lambda \geq vx_2 \right\}.$$

For this small problem the optimal solution can be determined without heavy computations. The optimal objective value for these problems is $-\frac{1}{\sqrt{2}}$ and is obtained for here-and-now decision $x_1^* = x_2^* = \frac{1}{2}$, and wait-and-see decision $y^* = \frac{1}{\sqrt{2}}$ in the primal formulation and $\lambda^* = \frac{1}{2}v$ in the dual formulation. The worst-case objective value in the dual formulation is achieved for $v^* = \frac{1}{\sqrt{2}}$. Suppose we solve the sampled version of the problem for only this worst-case scenario v^* . In that case, the sampled models looks like:

$$\min_{(x_1, x_2) \in \mathcal{X}, \bar{\lambda} \geq 0} \left\{ -\frac{2}{2\sqrt{2}} - \frac{1}{\sqrt{2}} + \bar{\lambda} \mid \bar{\lambda} \geq \frac{1}{\sqrt{2}}x_1, \bar{\lambda} \geq \frac{1}{\sqrt{2}}x_2 \right\},$$

which has optimal objective value of $-\frac{1}{\sqrt{2}}$. Hence, the lower bound that follows from the sampled version of the dual formulation is tight. If we now want to match a critical scenario ζ^* using (20) we get

$$\begin{aligned}\zeta^* &\in \arg \max_{\zeta \in \mathcal{U}} \{v^*(x_1^* \zeta_1 + x_2^* \zeta_2)\} \\ &= \arg \max_{\zeta \in \mathcal{U}} \{v^*(\frac{1}{2} \zeta_1 + \frac{1}{2} \zeta_2)\} \\ &= \{\zeta_1, \zeta_2 \geq 0 \mid \zeta_1 + \zeta_2 = 1\}.\end{aligned}$$

Notice that there is no unique maximizer to (20) for this problem. If we take extreme point $(\zeta_1^*, \zeta_2^*) = (1, 0)$, then the sampled version of the primal formulation is

$$\min_{(x_1, x_2) \in \mathcal{X}, \bar{y}} \{-\bar{y} \mid -1 + x_1 + \bar{y}^2 \leq 0\},$$

which has optimal objective value of -1 , which is strictly lower than the optimal solution to the original problem of $-\frac{1}{\sqrt{2}}$. Hence, in contrast to instances with only right-hand-side uncertainty, the sampled version of the dual formulation can give tighter lower bounds with left-hand side uncertainty.

C. Distribution on a network without commitments

Consider the linear variant of Problem (22) considered in §4, that is, distribution on a network *without* commitments, which can be written as the following two-stage robust linear optimization problem:

$$\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \inf_{y \geq 0} \left\{ \sum_{i=1}^N c_i x_i + \sum_{i,j=1}^N t_{ij} y_{ij} \mid \sum_{j=1}^N y_{ji} - \sum_{j=1}^N y_{ij} \geq \zeta_i - x_i \quad i = 1, \dots, N \right\}. \quad (27)$$

We choose $N \in \{10, 20, 30, 40, 50\}$ 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 10 for $i = 1, \dots, N$ stores. The individual maximum demand $\hat{\zeta}$ and the capacity K_i , $i = 1, \dots, N$, of each location is set to 20 units. The total demand in the network is set to $20\sqrt{N}$. Now that here we consider exact same problem setting as in Bertsimas and de Ruiter (2016), and compare the numerical performance of the lower bounding scheme proposed in §3.3 with the primal-dual lower bounding scheme proposed in Bertsimas and de Ruiter (2016). Table 6 reports the numerical results. One can observe that the optimality gaps obtained from our method (LB-P) almost halve the ones obtained from using the technique of Bertsimas and de Ruiter (2016) (LB-BR), where the optimality gap is computed via:

$$p\% = \frac{v(\text{D-L}) - v(\text{LB})}{v(\text{D-L})} \times 100\%,$$

where $v(\cdot)$ denotes the optimal value of the corresponding problems, *e.g.*, $v(\text{D-L})$ is the optimal value obtained from solving the dualized model of (27) with linear decision rules, while $\text{LB} \in \{\text{LB-P}, \text{LB-BR}\}$.

Table 6 Lot-sizing problem with $N \in \{10, 20, 30, 40, 50\}$. LB-P and LB-BR denotes the approximated optimality gap using the primal scenarios (see §3.3) and using the primal and dual scenarios in Bertsimas and de Ruiter (2016). All the numbers are the average of 10 randomly generated instances.

N	10	20	30	40	50
LB-P	6.0%	5.7%	6.0%	5.0%	5.2%
LB-BR	13.9%	12.9%	10.4%	11.2%	10.7%

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*, 52(1):189–209, 2018.
- A. Atamtürk and M. Zhang. Two-stage robust network flow and design under demand uncertainty. *Operations Research*, 55(4):662–673, 2007.
- J. Ayoub and M. Poss. Decomposition for adjustable robust linear optimization subject to uncertainty polytope. *Computational Management Science*, 13:216–239, 2016.
- 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. On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205, 2001.
- 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 I. Dunning. Multistage robust mixed integer optimization with adaptive partitions. *Operations Research*, 64(4):980–998, 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.
- D. Bertsimas, V. Goyal, and B. Y. Lu. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming*, 150(2):281–319, 2015.
- 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.
- X. Chen, M. Sim, P. Sun, and J. Zhang. A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357, 2008.
- V. Gabrel, M. Lacroix, C. Murat, and N. Remli. Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics*, 164(1):100–111, 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.

-
- G. Hanasusanto and D. Kuhn. Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. *Operations Research*, 66(3):849–869, 2018.
- G. Hanasusanto and G. Xu. Improved decision rule approximations for multi-stage robust optimization via copositive programming. *Optimization Online*, 2018. URL http://www.optimization-online.org/DB_FILE/2018/08/6776.pdf.
- D. Iancu, M. Sharma, and M. Sviridenko. Supermodularity and affine policies in dynamic robust optimization. *Operations Research*, 61(4):941–956, 2013.
- R. Jiang, J. Wang, M. Zhang, and Y. Guan. Two-stage minimax regret robust unit commitment. *IEEE Transactions on Power Systems*, 28(3):2271–2282, 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.
- A. Marandi and D. den Hertog. When are static and adjustable robust optimization problems with constraint-wise uncertainty equivalent? *Mathematical Programming*, 170(2):555–568, 2018.
- 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. Pinar and R. H. Tütüncü. Robust profit opportunities in risky financial portfolios. *Operations Research Letters*, 33(4):331–340, 2005.
- K. Postek and D. den Hertog. Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing*, 28(3):553–574, 2016.
- 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.
- 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.
- D. Simchi-Levi, H. Wang, and Y. Wei. Constraint generation for two-stage robust network flow problems. *INFORMS Journal on Optimization*, 1(1):49–70, 2018.
- 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.

- G. Xu and S. Burer. A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. *Computational Optimization and Applications*, 70(1):33–59, 2018. URL <https://arxiv.org/pdf/1609.07402v1.pdf>.
- B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41:457–461, 2013.
- J. Zhen and F. de Ruiter. Duality in adaptive robust linear optimization. *Working paper*, 2019.
- 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. 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*, 66(4):1086–1100, 2017b.
- J. Zhen, A. Marandi, D. den Hertog, and L. Vandenberghe. Disjoint bilinear programming: a two-stage robust optimization perspective. *Optimization Online*, 2018. URL http://www.optimization-online.org/DB_FILE/2018/06/6685.pdf.