A Primal-Dual Perspective on 
Adaptive Robust Linear Optimization

Jianzhe Zhen\textsuperscript{*1} and Frans de Ruiter\textsuperscript{2}

\textsuperscript{1}Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland
\textsuperscript{2}CQM, Eindhoven, The Netherlands

April 16, 2020

Abstract

Adaptive robust optimization is a modelling paradigm for multistage optimization under uncertainty where one seeks decisions that minimize the worst-case cost with respect to all possible scenarios in a prescribed uncertainty set. However, optimal policies for adaptive robust optimization models are difficult to compute. Therefore, one often restricts to the class of affine policies which give good solutions. We propose a simple set-lifting procedure to enhance the quality of affine policies for adaptive robust linear optimization models, which yields the optimal solution in finite iterations. Using a new lower bound technique, we compute tight progressive approximations on the optimal value and the obtained solution can be used to warm start the exact methods to reduce the computation time. Leveraging the primal and dual formulations, we improve the efficiency and effectiveness of the exact solution method of Zeng and Zhao (2013), and extend their method to solve models without complete recourse. The effectiveness of our method is demonstrated on a uncapacitated lot-sizing problem (with relatively complete recourse) and a capacitated lot-sizing problem (without relatively complete recourse).

Keywords: two-stage robust linear optimization, linear decision rules, reformulation linearization technique.

\textsuperscript{*}Corresponding author: jizhen@ethz.ch
1 Introduction

Both in theory and practice disregarding the uncertainties often results in suboptimal or even unimplementable decisions. Hence, there are many methods of dealing with uncertainty in real-life problems. One of the most well-known approaches is stochastic programming, which is very well applicable if the input parameters are random variables and the probability distributions are all known. Despite its wide popularity, stochastic programming is generally computationally demanding. For more details we refer the interested readers to Birge and Louveaux (2011). An alternative approach to deal with uncertainty is robust optimization. The primary motivation and goal of robust optimization is to provide a methodology to efficiently solve optimization models with uncertain parameters that does not require all distributional information to be known. In robust optimization one seeks for a fixed solution that minimizes the cost, and is immunized against any realization of the uncertain parameter in the uncertainty set. The techniques yield computationally tractable reformulations for a broad class of robust optimization models. The papers El Ghaoui and Lebret (1997), El Ghaoui et al. (1998) and Ben-Tal and Nemirovski (1998) are considered as the birth of this field. As shown in Ben-Tal and Nemirovski (2000), in many cases the robust objective value is only slightly higher than the version that neglects uncertainty.

In Ben-Tal et al. (2004) the classical robust optimization was extended to encompass adaptive robust optimization models. In contrast to robust optimization, some of the so-called wait-and-see decisions in dynamic models can be adapted at a later moment in time after (part of) the uncertain parameter has been revealed. An important special class of adaptive robust optimization models is disjoint bilinear programming (Zhen et al., 2018b). For instance, Coverage Set Algorithm for multi-agent systems as well as value function approximation problems can be reformulated as a disjoint bilinear program (Petrik and Zilberstein, 2007, 2011). Zhen et al. (2018b) reformulate bilinear models into adaptive robust optimization models, and establish the equivalence between affine policies and a well-known technique in global optimization, i.e., reformulation linearization technique (Sherali and Alameddine, 1992). Furthermore, the authors demonstrate the effectiveness and efficiency of robust optimization techniques on bimatrix games and concave quadratic minimization problems. For more applications, we refer to the book Ben-Tal et al. (2009).

Adaptive robust optimization yields less conservative decisions than robust optimization, but finding the optimal policies is a computationally intractable task. To circumvent these difficulties,
Ben-Tal et al. (2004) restrict the wait-and-see decisions to be affinely dependent on the uncertain parameters, an approach known as affine policies. This technique quickly became popular in practical models, see the survey by Yanikoglu et al. (2018). Bertsimas et al. (2010); Iancu et al. (2013) and Bertsimas and Goyal (2012) establish the optimality of affine policies for some specific classes of adaptive robust optimization models. Chen and Zhang (2009) further improve affine policies by extending the affine dependency to the auxiliary variables that are used in describing the uncertainty set. Henceforth, variants of piecewise affine decision rules have been proposed to improve the approximation while maintaining the tractability of adaptive robust optimization models. Such approaches include the deflected and segregated affine policies of Chen et al. (2008), the truncated affine policies of See and Sim (2009), and the bideflected and (generalized) segregated affine policies of Goh and Sim (2010). Bertsimas and de Ruiter (2016) have shown that two-stage robust linear optimization models admit an equivalent dual formulation. While the conservative approximations of the primal and dual formulations via affine policies are also equivalent, the sampled primal and dual models can be used to improve the progressive approximation scheme proposed by Hadjijyannis et al. (2011). In fact, affine policies were discussed in the early literature of stochastic programming but the technique had been abandoned due to suboptimality (Garstka and Wets, 1974). Interestingly, there is also a revival of using affine policies for adaptive stochastic optimization (Kuhn et al., 2011) after its effectiveness was shown in robust optimization approaches. Zhen et al. (2018a) characterize the structures of the optimal policies for a broad class of two-stage robust linear optimization models. However, even when the structure is known, constructing these optimal policies is hard. Ardestani-Jaafari and Delage (2016); Xu and Burer (2018); Hanasusanto and Kuhn (2018) reformulate two-stage robust linear optimization models with complete recourse into an equivalent copositive formulation, and use SDP approximation to compute conservative approximations. The complete recourse assumption is often made in two-stage robust optimization to ensure the second-stage model is feasible. Recently, many exact solution methods are developed to solve adaptive robust optimization models without complete recourse (Georghiou et al., 2017; Bertsimas and Shtern, 2018).

In Georghiou et al. (2017), extensive numerical experiments show that for two-stage robust linear models with complete recourse, the exact solution method of Zeng and Zhao (2013) outperforms the existing methods of Hadjijyannis et al. (2011); Kuhn et al. (2011); Zhen et al. (2018a); Georghiou
et al. (2017), while for models without complete recourse, the method of Zeng and Zhao (2013) is outperformed by all the other considered methods.

In this paper, we first improve quality of the affine policies and the existing progressive approximation scheme of Bertsimas and de Ruiter (2016). We then use the obtained solution from the progressive approximation to improve the efficiency of the exact solution method of Zeng and Zhao (2013). Furthermore, leveraging primal and dual formulations of Bertsimas and de Ruiter (2016), we show that models without complete recourse in its primal form often has a dual reformulation with complete recourse, which enables us to use the state-of-the-art method of Zeng and Zhao (2013) directly to solve the dual formulation. In particular, the main contributions of this paper are as follows.

1. We show that splitting equality constraints into two inequality constraints in the original primal formulation results in enhanced affine policies in the corresponding dual formulation, and the reformulation with inequalities corresponds to a lifted version of the original primal formulation popularized by Chen and Zhang (2009). We further extend our results to three-, four- and general multi-stage models, whereas previously only two-stage robust linear models could be dealt with by Bertsimas and de Ruiter (2016).

2. We show that further improvements of the affine policy can be deduced by lifting the uncertainty set in the primal formulation using the dual operation of the Fourier-Motzkin elimination procedure proposed by Zhen et al. (2018a).

3. We propose to adapt an efficient global optimization technique to compute progressive approximations for adaptive robust linear optimization models. These techniques lead to lower bounds of the optimal model that are much tighter than those presented in the literature.

4. Leveraging the primal and dual formulations, we extend the exact solution method of Zeng and Zhao (2013) to solve models without complete recourse, and demonstrate that the computational effort required to find optimal solutions by the exact solution method can be reduced significantly for both models with and without complete recourse by using our new lower bound solutions for a warm start. The obtained optimal solutions are then used to evaluate the effectiveness of our lower bound method on (capacitated) lot-sizing problems with distribution on a network.
Notation. Throughout the paper we write vectors and matrices in bold font and scalars in normal font. We use the vector $\mathbf{1}$ to denote the vector of all ones, and $[n]$, $n \in \mathbb{N}$, to denote the set of running indices $\{1, \ldots, n\}$. The vector $\mathbf{0}$ and matrix $O$ consist of only zero entries. All inequality signs represent componentwise inequalities. We use subscript $y_i$ to denote the $i$-th element of $y$. We use subscript $y^{(i)}$ and $y^{(j)}$ to distinguish different vectors within a set of $N$ vectors $(y^{(1)}, \ldots, y^{(N)})$.

2 Two-stage robust linear primal & dual models

We start with a model similar to those considered in Bertsimas and de Ruiter (2016):

$$\min_{x \in X} \left\{ c^\top x \mid \forall \zeta \in \mathcal{U} : \exists y : A(x)\zeta + By \geq h(x) \right\}, \quad (P)$$

where $X \subseteq \mathbb{R}^n_x$ represents the feasible set of the here-and-now decisions (some components of $x$ variables may be integer). In the first stage, the here-and-now decision $x$ is selected before the uncertain parameter $\zeta$ is revealed. The continuous wait-and-see decision $y \in \mathbb{R}^n_y$ is then chosen after $\zeta$ has been observed. The matrix $A(x) \in \mathbb{R}^{m \times n_\zeta}$ and the vector $h(x) \in \mathbb{R}^m$ depend affinely on $x$, that is,

$$A(x) = A_0 + \sum_{i=1}^{n_x} A_i x_i \quad \text{and} \quad h(x) = h_0 + \sum_{i=1}^{n_x} h_i x_i,$$

where $A_i \in \mathbb{R}^{m \times n_\zeta}$ and $h_i \in \mathbb{R}^m$ for all $i \in [n_x] \cup \{0\}$. Throughout this paper, we focus on models with fixed recourse, that is, we assume that the recourse matrix $B \in \mathbb{R}^{m \times n_y}$ is constant. Without loss of generality, there is no uncertainty in the objective function, i.e., the vector $c \in \mathbb{R}^{n_x}$ is constant, and the objective function only involves here-and-now decisions. More general objectives that accommodate uncertain parameters and wait-and-see decisions can always be reformulated as instances of $(P)$ by introducing an auxiliary epigraphical variable, see e.g., (Ben-Tal et al., 2009, p.10-11). We focus on nonempty polyhedral uncertainty sets of the form

$$\mathcal{U} = \{ \zeta : D\zeta \leq d \}, \quad (1)$$

where $D \in \mathbb{R}^{m_\zeta \times n_\zeta}$ and $d \in \mathbb{R}^{m_\zeta}$, which encapsulate the simplex, box, and budget uncertainty sets as special cases, see for example, Ben-Tal et al. (2004); Bertsimas and Sim (2004). Two-stage distributionally robust linear optimization models with polyhedral moments can be cast as a model in the form of $(P)$ (Bertsimas et al., 2017). A polyhedral ambiguity set can be used to model various...
uncertainties, such as polyhedral $\phi$-divergences, discrete Wasserstein ball, budget uncertainty sets for probabilities, 1- and $\infty$-norm balls.

**Remark 1 (Polyhedral approximation of a second-order cone).** Although restricting to polyhedral uncertainty sets may seem restrictive, note that we can make efficient polyhedral approximations for ellipsoidal uncertainty sets; see Ben-Tal and Nemirovski (2001). The resulting polyhedral uncertainty set is then, depending on the precision required, slightly larger than the original second-order-cone description. Lower precision levels give a larger uncertainty set, but are often acceptable as the uncertainty set is a choice made by the modeler.

By leveraging the strong duality for linear models, Bertsimas and de Ruiter (2016) consecutively dualize two-stage robust linear optimization model with respect to its wait-and-see decisions and the uncertain parameters to derive an equivalent dual formulation. After the first dualization step, they obtain an equivalent bilinear reformulation of (P).

**Lemma 1 (Bilinear reformulation).** Model (P) is equivalent to the disjoint bilinear model

$$
\min_{x \in X} \left\{ c^\top x \mid \forall (\zeta, w) \in \mathcal{U} \times \mathcal{W} : w^\top h(x) - \zeta^\top A(x)^\top w \leq 0 \right\},
$$

where $\mathcal{W} = \{w \in \mathbb{R}^m_+ : B^\top w = 0\}$.

**Proof.** It follows from Takeda et al. (2004) that we can write (P) as

$$
\min_{x \in X} \max_{\zeta \in \mathcal{U}} \min_{y} \left\{ c^\top x \mid A(x)\zeta + By \geq h(x) \right\}
$$

$$=
\min_{x \in X} \max_{\zeta \in \mathcal{U}} \max_{w \in \mathcal{W}} \left\{ c^\top x \mid w^\top h(x) - \zeta^\top A(x)^\top w \leq 0 \right\},
$$

where the equality follows from strong duality, which applies because $\{0\} \in \mathcal{W} \neq \emptyset$. $\square$

Note that the constraint (B) embeds a disjoint bilinear program. Therefore, model (B) is NP hard even if $x$ is fixed (Chen and Deng, 2006). This bilinear reformulation (B) is frequently used to solve two-stage robust linear optimization models in existing literature, e.g., Thiele et al. (2010); Gabrel et al. (2013); Zeng and Zhao (2013); Ardestani-Jaafari and Delage (2016); Simchi-Levi et al. (2018); Hanasusanto and Kuhn (2018); Xu and Burer (2018); Hanasusanto and Xu (2018); Bertsimas and Shtern (2018).

After the second dualization step, we recover the main result of Bertsimas and de Ruiter (2016).
Proposition 2 (Dual formulation (Bertsimas and de Ruiter, 2016)). Assume that polyhedral uncertainty set \( U \) as in (1) is nonempty. Then model (P) is equivalent to the two-stage robust linear optimization model

\[
\min_{x \in X} \left\{ c^\top x \mid \forall w \in \mathcal{W} : \exists \lambda \geq 0 : d^\top \lambda + h(x)^\top w \leq 0, \ D^\top \lambda = -A(x)^\top w \right\}.
\] (D)

Proof. By Lemma 1, we can write (B) as

\[
\min_{x \in X} \max_{w \in \mathcal{W}} \max_{\zeta \in \mathcal{U}} \left\{ c^\top x \mid w^\top h(x) - \zeta^\top A(x)^\top w \leq 0 \right\}
\]

\[
= \min_{x \in X} \max_{w \in \mathcal{W}} \min_{\lambda \in \mathbb{R}^m} \left\{ c^\top x \mid d^\top \lambda + h(x)^\top w \leq 0, \ D^\top \lambda = -A(x)^\top w \right\},
\]

where the equality follows from strong duality, which applies because \( U \) is nonempty. As shown in Takeda et al. (2004) the final line can be cast in the form of (D).

Note that the uncertain parameters and wait-and-see decisions in (P) are not restricted to be within the nonnegative orthant, which is different from the model considered in Bertsimas and de Ruiter (2016). Therefore, a slightly different dual formulation (D) is obtained, where (D) contains uncertain linear equality. In order to reestablish the equivalence between the affine policies for (P) and (D) as in the paper, we develop in Section 3.1 a new technique to deal with uncertain linear equalities in (D). Furthermore, we show via a primal-dual perspective in Section 3.4 that splitting each uncertain linear equality constraint into two inequalities in (D) is equivalent to imposing the extended affine policies of Chen and Zhang (2009) to the wait-and-see decisions in (P).

Finally, both (P) and (D) are two equivalent two-stage robust linear optimization models, while the dimension of wait-and-see decision in one formulation is equal to the number of constraints in the uncertainty set of its dual formulation. This “exchange” between the dimension of wait-and-see decision and the number of constraints in the uncertainty set result in a significant difference in the number of binary variables if the exact method of Zeng and Zhao (2013) is used to solve (P) and (D); see Section 3.5. In Section 5.1, we solve the primal and dual formulations of a lot-sizing problem via the exact method of Zeng and Zhao (2013), and demonstrate that the dual formulation is much easier to solve than the primal formulation due to the significant reduction in the number of binary variables.
3 A primal-dual perspective on existing methods

3.1 Affine policies

Adaptive robust optimization models are in general computationally intractable. To circumvent the intractability, Ben-Tal et al. (2004) propose a tractable conservative approximation by restricting to affine policies. In this section, we show that imposing affine policies to the wait-and-see variables in either (P) or (D) lead to equivalent conservative approximations, while the resulting descriptions of the former involve fewer variables than the latter. We extend our discussion to the multi-stage case, whereas previously only two-stage models were considered in Bertsimas and de Ruiter (2016).

Without loss of generality, we fix \( x \in X \) and focus on the constraints of primal and dual formulations as both formulations share the same objective function. Restricting the wait-and-see decisions in (P) to affine policy

\[
y(\zeta) = y_0 + Y\zeta, \tag{2}
\]

where \( y_0 \in \mathbb{R}^n_x \) and \( Y \in \mathbb{R}^{n_x \times n_\zeta} \), we find a conservative approximation of the feasible set of (P),

\[
\forall \zeta \in \mathcal{U} : \begin{cases}
A(x)\zeta + B(y_0 + Y\zeta) \geq h(x) \\
y_0, Y \text{ free}
\end{cases} \iff \begin{cases}
By_0 \geq h(x) + \Pi^\top d \\
-BY = \Pi^\top D + A(x) \\
\Pi \in \mathbb{R}_{++}^{m_\zeta \times m}, \ y_0, Y \text{ free.}
\end{cases} \tag{3}
\]

The “\( \iff \)” follows from the standard robust counterpart techniques to get rid of the “\( \forall \)” quantifier, see Ben-Tal et al. (2009).

Similarly, restricting the wait-and-see decisions in (D) to the following affine policy

\[
\lambda(w) = \Lambda w, \tag{4}
\]

where \( \Lambda \in \mathbb{R}^{m_\zeta \times m} \), we find a conservative approximation of the feasible set of (D),

\[
\forall w \in \mathcal{W} : \begin{cases}
d^\top \Lambda w + h(x)^\top w \leq 0 \\
D^\top \Lambda w = -A(x)^\top w \\
\Lambda w \geq 0 \\
\Lambda \text{ free}
\end{cases} \iff \begin{cases}
Bv \geq h(x) + \Lambda^\top d \\
-BV = \Lambda^\top D + A(x) \\
B\Omega \geq -\Lambda^\top \\
v, V, \Lambda, \Omega \text{ free.}
\end{cases} \tag{5}
\]
where $V \in \mathbb{R}^{n_y \times n_z}$, $\Lambda \in \mathbb{R}^{m_z \times m}$, and $\Omega \in \mathbb{R}^{n_y \times m_z}$. Note that we did not include an intercept $\lambda_0$ in (4), because from model (D) we can see that for the scenario $w = 0 \in \mathcal{W}$ we need to have $\lambda(0) = \lambda_0 + \Lambda 0 = 0$, which implies $\lambda_0 = 0$. Note that the left-hand-side of (5) contains uncertain linear equality constraints. We propose a novel approach to deal with uncertain linear equality and use it to derive the right-hand-side of (5), i.e., the “$\iff$” in (5) follows from the standard robust counterpart techniques and the following lemma.

**Lemma 3** (Uncertain linear equality). The following statement holds.

$$ \forall w \in \mathcal{W} : \alpha^\top w = 0 \iff \exists \eta : B\eta = \alpha. $$

_Proof._ “$\iff$”: Assume there exists a $\eta$ such that $B\eta = \alpha$ holds, then we find $$ \forall w \in \mathcal{W} : \eta^\top B^\top w = 0. $$

“$\implies$”: Assume now that for all $w$ in $\mathcal{W}$ the equality $\alpha^\top w = 0$ holds, then we find

$$ 
\min_x \{0 \mid \forall w \in \mathcal{W} : \alpha^\top w = 0\} = \min_x \max_{\nu_0} \min_{\nu_0} \max_{\eta} \{0 \mid B\eta = \nu_0\alpha\} 
$$

where the first equality follows from Lagrangian duality and Lemma 1. The last equality holds due to the homogeneity of linear systems. 

The tractable counterparts in (3) and (5) are intimately related to each other.

**Proposition 4** (Equivalence of primal and dual affine policies (Bertsimas and de Ruiter, 2016)).

The solution $(x, y_0, Y, \Pi)$ is feasible in (3) if and only if $(x, \Lambda, v, V, \Omega)$ is feasible in (5), where

$$ y_0 = v + \Omega d, \quad Y = V - \Omega D, \quad \Pi^\top = B\Omega + \Lambda^\top. $$

Proposition 4 shows that one can construct a feasible policy to (3) from a feasible policy to (5) with the same objective value, and vice versa. Despite sets (3) and (5) have exactly the same number of (in)equality constraints, the set (5) always involves more optimization variables (because of $\Omega$) than those in (3). Thus, it is always better to consider (3) instead of (5) because $\Omega$ is a redundant variable in (5), that is, setting $\Omega = O$ in (5) would result in (3). On the other hand, for a variant
formulation of (P) with non-negative uncertain parameters and wait-and-see decisions, Bertsimas and de Ruiter (2016) show both approximations of primal and dual formulations via affine policies also have the same number of optimization variables and constraints, however, since the numbers of affine constraints and sign constraints of both approximations may differ, the approximation with fewer affine constraints is much faster to solve via existing solvers.

**Remark 2** (Static policies for primal and dual formulations). It may be surprising that static policies in primal and dual formulation, i.e., policies with \( y(\zeta) = y_0 \) and \( \lambda(w) = \lambda_0 \) for some \( y_0 \) and \( \lambda_0 \), are not always equivalent. To see this, consider the follow instance of (P) and its dual formulation:

\[
\min_{x} \max_{\zeta \in \mathcal{U}} \min_{y \in \mathbb{R}_+^n} \{ 1^T y \mid y \geq \zeta \} \quad \iff \quad \min_{x} \max_{w \in [0,1]^n} \min_{\lambda \in \mathbb{R}_+} \{ \lambda \mid \lambda 1 \geq w \},
\]

where \( \mathcal{U} = \{ \zeta \in \mathbb{R}_+^n \mid 1^T \zeta \leq 1 \} \) is a simplex set. If static policies are imposed to the wait-and-see decisions \( y \) and \( \lambda \), then the corresponding optimal objective values are \( n \) for the primal and \( 1 \) for the dual formulation. Moreover, if the following slightly modified version of the model is considered:

\[
\min_{x} \max_{\zeta \in \mathcal{U}} \min_{y \in \mathbb{R}_+^n, 1^T y \leq x} \{ x \mid y \geq \zeta, 1^T y \leq x \} \quad \iff \quad \min_{x} \max_{w \in [0,\tau]^n} \min_{\lambda \in \mathbb{R}_+} \{ x \mid \lambda 1 \geq w, x \tau \geq \lambda \},
\]

and static policies are imposed to the wait-and-see decisions \( y \) and \( \lambda \), then the corresponding optimal objective value is still \( n \) for the primal, whereas the dual formulation with static policies is infeasible.

### 3.2 Multistage adaptive robust models

So far in the discussion above, as well as in Bertsimas and de Ruiter (2016), we have only considered two-stage models. The difficulty with multistage models lies in non-anticipativity of the policies. That is, a wait-and-see decision that has to materialize in period \( i \) is only allowed to use information on the uncertain parameter that is available up to period \( i \). In primal models the nonanticipativity rules have been introduced in Ben-Tal et al. (2004). They only allow \( Y_{ij} \) to be nonzero whenever the value of \( \zeta_j \) is known when the value of \( y_i(\zeta) \) is known for the \( j \)-th wait-and-see decision. However, if these nonanticipativity rules are known, we can use the result from proposition 4 to formulate conditions on \( V \) and \( \Omega \). Namely, for some \( i \) and \( j \) the restriction

\[ Y_{ij} = 0, \]
which implies that information on parameter $j$ is not known for wait-and-see decision $i$ is to adding constraint

$$V_{ij} - \sum_{k=1}^{m_c} \Omega_{ik} D_{kj} = 0$$

to model (D). Therefore, we can also incorporate nonanticipativity rules in the dual model once the restrictions on the primal model are known.

### 3.3 Reformulation linearization technique (Sherali and Alameddine, 1992)

Ardestani-Jaafari and Delage (2016) show that the approximation from affine policies for two-stage robust linear models is equivalent to the approximation of the bilinear reformulation via reformulation linearization techniques of Sherali and Alameddine (1992). Dualizing with respect to the intercept $y_0$ and the coefficient matrix $Y$ of the affine policy in the left hand side of (3) using the “primal-worst equals dual-best” technique of Beck and Ben-Tal (2009), we find

$$\forall \left\{ w, \{w^{(j)} \}_{j \in [n_c]} \right\} \in \mathcal{V}': h^T(x)w - \sum_{j=1}^{n_c} A_j^T(x)w^{(j)} \zeta_j \leq 0$$

(6)

where $\mathcal{V}'$ is a bilinear uncertainty set defined through

$$\mathcal{V}' = \left\{ \left( w, \{w^{(j)} \}_{j \in [n_c]} \right) \Bigg| B^T w = 0, B^T w^{(j)} = 0, j \in [n_c], \sum_{j=1}^{n_c} D_j^T \zeta_j w_i^{(j)} \leq dw_i, i \in [m] \right\}.$$

Similarly, dualizing with respect to $\Lambda$ in the left hand side of (5), and eliminate some redundant variables/constraints, we find

$$\forall \left( \{w, \zeta^{(i)} \}_{i \in [m]} \right) \in \mathcal{V}: \sum_{i=1}^{m} w_i h_i(x) - \sum_{i=1}^{m} A_i^T(x) \zeta^{(i)} w_i \leq 0,$$

(7)

where $\mathcal{V}$ is a bilinear uncertainty set defined through

$$\mathcal{V} = \left\{ \left( \{w, \zeta^{(i)} \}_{i \in [m]} \right) \Bigg| B^T w = 0, D \zeta^{(i)} \leq d, i \in [m], \sum_{i=1}^{m} B_i^T w_i \zeta^{(i)} = 0 : Y_j, j \in [n_c] \right\},$$

and $Y_j, j \in [n_c]$, are the dual variables of the corresponding constraints. Substituting respectively $\Gamma_i : \zeta^{(i)} w_i, i \in [m]$ and $\Gamma_j : w^{(j)} \zeta_j, j \in [n_c]$, in (7) and (6), respectively, we have

$$\forall (w, \Gamma) \in \mathcal{Y}: h^T(x)w - \sum_{i=1}^{m} A_i^T(x) \Gamma_i \leq 0,$$

(8)
where $\Upsilon$ is the linearized uncertainty set defined through

$$\Upsilon = \{(w, \Gamma) \mid B^\top w = 0, B^\top \Gamma_j = 0, j \in [n_c], D \Gamma_i \leq dw_i, i \in [m]\}.$$ 

One can use the result of Gorissen et al. (2014, Lemma 1) to show the equivalence of (6), (7) and (8), which applies because $U$ is compact. This confirms that the approximations of (P) and (D) via affine policies are of same quality; also see Proposition 4.

**Remark 3** (Non-anticipative constraints using reformulation linearization technique). In order to take into account the non-anticipative constraints using reformulation linearisation techniques, one can simply omit the constraints that are corresponding to the 0 entries of $Y$ in (7), which leads to ‘fewer constraints’ in the uncertainty sets $V$ and $\Upsilon$.

### 3.4 Extended affine policies (Chen and Zhang, 2009)

We first show that the conservative approximations from affine policies can be improved by simply splitting each equality constraint into two inequality constraints. Furthermore, we establish the equivalence of this “splitting” technique and the existing robust optimization technique via the developed primal-dual scheme.

Consider a variant formulation of (D), where the equality constraint is splitted into two inequality constraints, that is,

$$\forall w \in \mathcal{W} : \exists \lambda \geq 0 : \begin{cases} d^\top \lambda + h(x)^\top w & \leq 0 \\
 D^\top \lambda & \geq -A(x)^\top w \\
 D^\top \lambda & \leq -A(x)^\top w. \end{cases}$$

(9)

Such a splitting does not affect the feasible region of (D). However, it maybe surprising that by imposing the affine policy in (4) to the second-stage variable $\lambda$ in (9) leads to a tighter approximation

$$\begin{cases} Bv & \geq h(x) + \Lambda^\top d \\
 -BV^+ & \geq \Lambda^\top D + A(x) \\
 BV^- & \geq -\Lambda^\top D - A(x) \\
 B\Omega & \geq -\Lambda^\top \\
 v, V^+, V^-, \Lambda, \Omega & \text{free,} \end{cases}$$

(10)
where $V^+, V^- \in \mathbb{R}^{ny \times n\zeta}$. The set (5) constitutes a subset of (10), and two sets coincide if and only if $V^+ - V^- = O$. Applying Proposition 2, we obtain the corresponding primal formulation of (9),

$$\forall (\zeta^+, \zeta^-) \in U^+: \exists y: A(x) (\zeta^+ - \zeta^-) + By \geq h(x),$$  \hspace{1cm} (11)

where the uncertainty set is defined through $U^+ = \{(\zeta^+, \zeta^-) \in \mathbb{R}^{n\zeta}_+ \times \mathbb{R}^{n\zeta}_+ : D(\zeta^+ - \zeta^-) \leq d\}$. Therefore, splitting the uncertain equality constraint leads to segregating the free uncertain parameters in its dual formulation. The affine policy (2) for (P) can be readily extended by incorporating the extra auxiliary variable $\zeta^-$, that is,

$$y(\zeta^+, \zeta^-) = y_0 + Y^+ \zeta^+ - Y^- \zeta^-,$$

where $Y^+, Y^- \in \mathbb{R}^{ny \times n\zeta}$. Then, we recover the so-called extended affinely adaptive robust counterpart of Chen and Zhang (2009)

$$\begin{aligned}
B y_0 & \geq h(x) + \Pi^T d \\
BY^+ & \geq -\Pi^T D - A(x) \\
-BY^- & \geq \Pi^T D + A(x) \\
\Pi & \in \mathbb{R}^{m\zeta \times m}_+, y_0, Y^+, Y^- \text{ free.}
\end{aligned}$$  \hspace{1cm} (12)

The set (3) is a subset of (12), and two sets coincide if and only if $Y^+ - Y^- = O$. Similarly as in Proposition (4), one can establish the following relationship between (10) and (12), that is, $y_0 = v + \Omega d$, $Y^+ = V^- - \Omega D$, $Y^- = V^+ - \Omega D$ and $\Pi^T = B\Omega + \Lambda^T$, which indicates that it is better to consider (12) instead of (10) because the variable $\Omega$ is redundant in (10).

Consider a variant formulation of (P), where the wait-and-see decisions are segregated as follows,

$$\forall \zeta \in U : \exists y^+, y^- \geq 0: A(x)\zeta + B(y^+ - y^-) \geq h(x),$$  \hspace{1cm} (13)

where $y^+, y^- \in \mathbb{R}^{ny}$. Imposing affine policy (2) to both $y^+, y^- \in \mathbb{R}^{ny}_+$, we have

$$\begin{aligned}
B (y^+_0 - y^-_0) & \geq h(x) + \Pi^T d \\
B (Y^+ - Y^-) & \geq -\Pi^T D - A(x) \\
y^+_0 & \geq (\Xi^+)^T d, \quad D^T \Xi^+ = -(Y^+)^T \\
y^-_0 & \geq (\Xi^-)^T d, \quad D^T \Xi^- = -(Y^-)^T \\
\Pi & \in \mathbb{R}^{m\zeta \times m}_+, \Xi^+ \in \mathbb{R}^{m\zeta \times n\zeta}_+, \Xi^- \in \mathbb{R}^{m\zeta \times n\zeta}_+ \\
y^+_0, y^-_0, Y^+, Y^- \text{ free.}
\end{aligned}$$  \hspace{1cm} (14)
Note that whenever the solution \((x, \Pi, y_0^+, Y^+, Y^-, \Xi^+, \Xi^-)\) is feasible in (14), one can construct a solution \((x, \Pi, y_0^+ - y_0^-, Y^+ - Y^-)\) that is feasible in (3), which implies segregating the wait-and-see decisions increases the computational complexity and may lead to solutions with worse (higher) objective values.

**Remark 4** (Modelling guide). *Instead of imposing affine policies to the wait-and-see variables directly, it is better to split equalities with two inequalities, and segregate the free uncertain parameters and then impose the extended affine policies with extra auxiliary variables. On the other hand, it is better not to segregate the free wait-and-see decision variables.*

### 3.5 Exact solution method (Zeng and Zhao, 2013)

In this subsection, we discuss the differences in solving the primal and dual formulations using the column-and-constraint generation (CCG) method of Zeng and Zhao (2013). To this end, for ease of exposition, consider the following variant of \((P)\)

\[
\min_{x \in X} \max_{\zeta \in U'} \min_{y \geq 0} \left\{ c^T x + b^T y \mid A(x)\zeta + By \geq h(x) \right\}, \tag{P'}
\]

where \(b \in \mathbb{R}^n_{+}\) and \(U' = \{ \zeta \in \mathbb{R}_+^n : D\zeta \leq d \}\) is nonempty. The model \((P')\) is feasible if there exists a \(x \in X\) such that for any \(\zeta \in U'\) the second-stage model is feasible, and the feasibility of \((P')\) is guaranteed if it has complete recourse.

**Definition 1** (Complete recourse). *We say that \((P')\) has complete recourse if there exists a \(y \in \mathbb{R}^n_{+}\) with \(By > 0\).*

Typically, a weaker condition is assumed in stochastic programming to ensure that the second-stage model of \((P')\) is feasible.

**Definition 2** (Relatively complete recourse). *We say that \((P')\) has relatively complete recourse if for any fixed \((x, \zeta) \in X \times U'\) the second-stage model of \((P')\) is feasible.*

One can show that \((P')\) admits the following dual formulation

\[
\min_{x \in X} \max_{w \in W'} \min_{\lambda \geq 0} \left\{ c^T x + w^T h(x) + d^T \lambda \mid D^T \lambda \geq -A^T(x)w \right\}, \tag{D'}
\]

where \(W' = \{ w \in \mathbb{R}^m_+ : B^T w \leq b \}\). Since the proof for the equivalence of \((P')\) and \((D')\) is almost identical to the proofs in Section 2, we relegate the proof in Appendix A. Similarly, \((D')\) has complete recourse if there exists \(\lambda \in \mathbb{R}^m_{+}\) with \(D^T \lambda > 0\).
Remark 5 (Complete recourse for primal and dual formulations). Note that it is possible that the primal formulation has (relatively) complete recourse while the dual formulation does not, or vice versa. If (P') has a box or budgeted uncertainty set, then (D') always has complete recourse.

Let \( \hat{x} \in X \) be the starting infeasible solution of (P'), for instance, from the progressive solution method in Section 4.2. If (P') or (D') has (relatively) complete recourse, then it follows from Zeng and Zhao (2013) that one can solve subproblems, which are mixed binary linear programs with \((m + n_y)\) or \((n_\zeta + m_\zeta)\) binary variables, respectively, to determine valid cuts for the given \( \hat{x} \). If a valid cut is found, then this cut is used to update the solution \( \hat{x} \). This procedure can be repeated till no valid cut can be found, and the obtained solution \( \hat{x}' \) is optimal for (P').

As long as one of the formulations (P') and (D') has (relatively) complete recourse, then the CCG method of Zeng and Zhao (2013) can be applied. In Section 5.2, we consider a two-stage robust linear model without (relatively) complete recourse, while its dual formulation has complete recourse, and we solve the dual formulation using the CCG method of Zeng and Zhao (2013). Moreover, even if both (P') and (D') have (relatively) complete recourse, the number of binary variables in the corresponding subproblems could be significantly different. For instance, for the model considered in Section 5.2, the number of binary variables in the corresponding subproblems for (P') and (D') are \((n_x^2 + n_x)\) and \((2n_x + 1)\), respectively, while for the model considered in Section 5.1, the number of binary variables in the corresponding subproblems for (P') and (D') are \((2n_x^2 + n_x)\) and \((2n_x + 1)\), respectively.

Remark 6 (Extensions to the CCG method of Zeng and Zhao (2013)). The CCG method of Zeng and Zhao (2013) can also be used to solve (P') with a general convex \( U' \) and (D') with a general convex \( W' \), and the corresponding subproblems for (P') and (D') then become mixed binary nonlinear models.

4 Novel conservative and progressive solution methods

In this section, we first propose a novel systemic lifting procedure for compact polyhedral sets to enhance affine policies, which provides conservative approximations of (P). Next, we construct two algorithms which are adapted from a iterative solution method in global optimization to compute tight progressive approximations for (P).
4.1 A new conservative solution method

In Section 3.4, we have shown that segregating the uncertain parameter into two pieces improves the approximation from affine policies. Other set lifting techniques are considered in Chen et al. (2007, 2008) and Goh and Sim (2010) to enhance the approximation from affine policies for multi-stage stochastic programming models. Georghiou et al. (2017) propose a lifting procedure using mixed integer programming for two-stage robust linear optimization models with polyhedral uncertainty sets. Alternatively, we propose a generic and systematic lifting procedure via Algorithm 1 based on the “dual” of Fourier-Motzkin elimination. More specifically, we apply Algorithm 1 to lift the polyhedral set \( \mathcal{Z} \) into a higher dimension, where \( \mathcal{Z} \) is defined through

\[
\mathcal{Z} = \{ z \in \mathbb{R}_+^{n_x} \mid \mathbf{p}_i \top z = q_i, \ i \in [m] \},
\]

where \( \mathbf{p}_i \in \mathbb{R}^{n_x} \) and \( q_i \in \mathbb{R} \) for every \( i \in [m] \). We illustrate Algorithm 1 via Example 1, and relegate the formal description to Appendix B.

**Example 1** (Lifting a budgeted set). Consider a budgeted set with a given budget \( \tau \in \mathbb{R}_+ \), then

\[
\mathcal{Z} = \{ z \in \mathbb{R}_+^{n_x} \mid z \leq 1, \ 1 \top z \leq \tau \}
\]

\[
\iff \quad \mathcal{Z} = \{ z \in \mathbb{R}_+^{n_x} \mid \exists \mathbf{v} \in \mathbb{R}_+^{n_x} : z + \mathbf{v} = 1, \ 1 \top z \leq \tau \}
\]

\[
\iff \quad \mathcal{Z} = \left\{ z \in \mathbb{R}_+^{n_x} \mid \exists \mathbf{v} \in \mathbb{R}_+^{n_x} : z_1 + v_1 - z_2 - v_2 = 0, \ z_i + v_i = 1, \ i \in [n_z] \right\}.
\]

Replacing \( z_1 \leftarrow w_{11} + w_{12}, \ v_1 \leftarrow w_{21} + w_{22}, \ z_2 \leftarrow w_{11} + w_{21} \) and \( v_2 \leftarrow w_{12} + w_{22} \) in \( \mathcal{Z} \), by construction we have

\[
w_{11} + w_{12} + w_{21} + w_{22} - w_{11} - w_{21} - w_{12} - w_{22} = z_1 + v_1 - z_2 - v_2 = 0,
\]
hence, the constraint $z_1 + v_1 - z_2 - v_2 = 0$ can be removed in $Z$, and then we have

$$
Z' = \begin{cases}
(z_3, \ldots, z_{n_z}, w_{11}, w_{21}, w_{22}) \in \mathbb{R}^{n_z+1}_+ & \exists \exists w_{22} \geq 0, \ z_i + v_i = 1, \ i \in I \\
& w_{11} + w_{21} + w_{12} + w_{22} = 1 \\
& \exists v_i \geq 0, \ i \in I \\
& 2w_{11} + w_{12} + w_{21} + \sum_{i=3}^{n_z} z_i \leq \tau
\end{cases}
$$

where $I = \{3, \ldots, n_z\}$. By eliminating one constraint, the set $Z$ is lifted into $\mathbb{R}^{n_z+1}$.

After each iteration of Algorithm 1, some new variables are introduced in the lifted set, which can be used to enhance the affine policy to improve the conservative approximation. One can of course iteratively eliminate all but one constraint in $U$ of (P). The resulting lifted set is a simplex, and affine policies are optimal (Bertsimas and Goyal, 2012) for two-stage models. Algorithm 1 provides a systematic way to solve (P) in finite iterations, and can easily incorporate non-anticipative constraints using affine policies to solve multistage models. Furthermore, since the lifting procedure only affects the description of the uncertainty set, it can be used as preprocessor procedure to complement any existing methods for adaptive robust linear optimization models with polyhedral uncertainty sets.

Remark 7 (Primal and dual of Fourier-Motzkin elimination). From a primal-dual perspective, lifting the uncertainty set $U$ in (P) would increase the dimension of the uncertain parameter and the number of constraints in (D), and reduce the number of constraints in $U$ and wait-and-see variables in (D). In other words, lifting the uncertainty set $U$ in (P) is equivalent to eliminating the corresponding wait-and-see decision variables in (D) via Fourier-Motzkin elimination. Due to this close relationship between primal and dual operations of Fourier-Motzkin elimination, we refer interested readers to Zhen et al. (2018a) for extensive numerical experiments on the performance of Fourier-Motzkin elimination for adaptive robust optimization models.

4.2 A new progressive solution method

One direct way of approximately solving adaptive robust optimization models is to consider a finite subset of scenarios sampled from the uncertainty set where each scenario with its own wait-and-see
decision, instead of determining complex (maybe non-convex) decision rules that are feasible for all possible realizations of the uncertain parameters. The sampled version of the primal model (P) is a linear optimization model

\[
\min_{x \in X} \left\{ c^\top x \mid \exists y^{(k)}, k \in [K]: A(x)\zeta^{(k)} + By^{(k)} \geq h(x), k \in [K] \right\},
\]

(P-S)

where \( \{\zeta^{(1)}, \ldots, \zeta^{(K)}\} = \overline{U} \subset U \). The minimum of (P-S) provides a lower bound on that of (P).

The optimal here-and-now decision \( \hat{x} \) for (P-S) is only feasible for the \( K \) scenarios sampled from \( U \), and there could be scenarios in \( U \) for which a higher minimum is attained, which makes \( \hat{x} \) infeasible.

Similarly, the sampled version of the dual formulation (D) is a linear optimization model

\[
\min_{x \in X} \left\{ c^\top x \mid \exists \lambda^{(k)} \geq 0, k \in [K]: d^\top \lambda^{(k)} + h(x)^\top w^{(k)} \leq 0, k \in [K] \right\},
\]

(D-S)

where \( \{w^{(1)}, \ldots, w^{(K)}\} = \overline{W} \subset W \).

The question that remains is how to choose the finite set of scenarios. One way to do this would be to include all extreme points. In that case, Bemporad et al. (2003) prove that the sampled versions are in fact optimal. Another way to obtain a small and effective finite set of scenarios for two-stage robust linear models is described by Hadjiyiannis et al. (2011), which takes scenarios that are binding for the model solved with affine policies, hoping that the same set of scenarios is also binding for the optimal policies. 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. Bertsimas and de Ruiter (2016) enhance the approach of Hadjiyiannis et al. (2011) by considering the binding scenarios from the approximations of both primal and dual formulations via affine policies, and combining the corresponding constraints of (P-S) and (D-S) to compute bounds on the optimal values for two-stage robust linear optimization models with polyhedral uncertainty sets, e.g.,

\[
\min_{x \in X} \left\{ c^\top x \mid \exists \lambda^{(k)} \geq 0, y^{(k)}, k \in [K]: A(x)\zeta^{(k)} + By^{(k)} \geq h(x), k \in [K] \right\}
\]

(PD-S)

Alternatively, in order to obtain good primal and dual scenarios for the scenario model (PD-S), one can solve the following static robust optimization model obtained from the reformulation lin-
earization technique (see Section 3.3)

$$\min_{x \in \mathcal{X}} \left\{ c^\top x \mid \forall (w, \Gamma) \in \Upsilon : h^\top(x)w - \sum_{i=1}^{m} A_i^\top(x)\Gamma_i \leq 0 \right\},$$

where the linearized uncertainty set $\Upsilon$ is defined as in (8). Note that each row and column of $\Gamma$ correspond to a primal and a dual scenario, respectively, which can be used as candidate scenarios for the progressive solution methods.

We propose a solution scheme that progressively solves two-stage robust linear models. Consider an optimal here-and-now decision $\hat{x}$ of (P-S) given $\overline{U} \subset U$, which may be infeasible for some scenarios in $U$. Given a scenario $\hat{\zeta}$ for (P-S) one can determine a scenario for (D-S) by solving a linear optimization model

$$\hat{w} = \arg \max_{w \in W} \left\{ w^\top h(\hat{x}) - w^\top A(\hat{x})\hat{\zeta} \right\}, \quad (15)$$

and the obtained dual scenario $\hat{w}$ constitutes a valid cut for the scenario model (D-S) if the corresponding objective value $v_{\hat{w}}$ of (15) is strictly larger than 0. Likewise, given a scenario $\hat{w}$ for (D-S) one can construct a scenario by solving the following linear optimization model

$$\hat{\zeta} = \arg \max_{\zeta \in U} \left\{ \hat{w}^\top h(\hat{x}) - \zeta^\top A(\hat{x})^\top \hat{w} \right\}, \quad (16)$$

and the obtained primal scenario $\hat{\zeta}$ constitutes a valid cut for the scenario model (P-S) if the corresponding objective value $v_{\hat{\zeta}}$ of (16) is strictly larger than 0. Therefore, one can iteratively solve (15) and (16) to generate valid cuts for (P-S) and (D-S). In Appendix C, Algorithm 2 is formally described. This iterative procedure is inspired by the mountain climbing method for bilinear models (Nahapetyan, 2008), and a similar algorithm is also considered in Bertsimas and Shtern (2018).

An important special case of (P) is when $A(\hat{x}) = A$, that is, (P) has right-hand-side uncertainties. A priori optimality bounds for the conservative approximation of two-stage robust linear models with right-hand-side uncertainties via affine policies are provided in Bertsimas and Goyal (2012), Bertsimas and Bidkhori (2015) and El Housni and Goyal (2018).

In the remainder of this section, we improve the efficiency of Algorithm 2 for the case when $A(\hat{x}) = A$ in (P). To this end, we first show that the primal scenario $\hat{\zeta}$ from (16) for a given $\hat{w} \in W$ produces a tighter approximation than that of $\hat{w}$.
Proposition 5 (Dominating scenarios). If \( A(x) = A \) in (P), then the primal scenario \( \hat{\zeta} \) that is obtained from (16) given a dual scenario \( \hat{w} \in W \) satisfies

\[
\left\{ x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m : A\hat{\zeta} + By \geq h(x) \right\} \subseteq \left\{ x \in \mathbb{R}^n \mid \exists \lambda \in \mathbb{R}^m_+ : D^\top \lambda = -A^\top \hat{w} \right\}.
\]

Proof. Given a dual scenario \( \hat{w} \in W \). Dualize the following set with respect to \( \lambda \), we find

\[
\left\{ x \in \mathbb{R}^n \mid \exists \lambda \in \mathbb{R}^m_+ : d^\top \lambda + h(x)^\top \hat{w} \leq 0 \right\} = \left\{ x \mid \forall \zeta \in U : \hat{w}^\top h(x) - \zeta^\top A^\top \hat{w} \leq 0 \right\}
\]

\[
= \left\{ x \mid \hat{w}^\top h(x) - \min_{\zeta \in U} \zeta^\top A^\top \hat{w} \leq 0 \right\}
\]

\[
= \left\{ x \mid \hat{w}^\top h(x) - \hat{\zeta}^\top A^\top \hat{w} \leq 0 \right\}
\]

\[
\supseteq \left\{ x \mid \forall w \in W : w^\top h(x) - \hat{\zeta}^\top A^\top w \leq 0 \right\}
\]

\[
= \left\{ x \in \mathbb{R}^n \mid \exists y \in \mathbb{R}^m : A\hat{\zeta} + By \geq h(x) \right\}.
\]

The first equality is due to strong duality of linear programming, while the third equality holds because \( A \) does not depend on \( x \). The “\( \supseteq \)” follows from the fact that \( \hat{w} \in W \). The last equality is again due to strong duality of linear programming.

A more general version of Proposition 5 for two-stage robust nonlinear optimization models was introduced in de Ruiter et al. (2017); Zhen et al. (2017). It follows from Proposition 5 that the dual scenarios in Algorithm 2 are dominated by the corresponding primal scenarios, and hence, can be omitted. Therefore, we introduce an improved version of Algorithm 2 in Appendix C, that is, Algorithm 3, to solve (P) with \( A(x) = A \). It turns out that the first iteration of Algorithm 3 coincides with the progressive method of de Ruiter et al. (2017); Zhen et al. (2018b). We show via numerical experiments that the solutions from Algorithm 3 can be computed efficiently and provide tight lower bounds that are much tighter than the lower bounds from the existing methods. Furthermore, in Section 5.1, we use the solution obtained from Algorithm 3 to warm start the exact solution method of Zeng and Zhao (2013) and consequently reduce the computational time.

5 Lot-sizing with distribution on a Network

The model setting here is adopted from Bertsimas and de Ruiter (2016). In the lot-sizing problem we determine the stock allocation \( x_i \) for \( i \in [n_x] \) stores prior to knowing the realization of the
demand at each location. The capacity of the stores is incorporated in $X$. The demand $ζ$ is uncertain. After we observe the realization of the demand we can transport stock $y_{ij}$ from store $i$ to store $j$ at unit cost $t_{ij}$ to meet all demand. The aim is to minimize the worst case storage costs (with unit costs $c_i$) and the cost arising from shifting the products from one store to another. The network flow model can expressed as a two-stage robust linear optimization model

$$
\min_{x \in X} \max_{\zeta \in U} \min_{y \geq 0} \left\{ c^\top x + \sum_{i,j \in [nx]} t_{ij} y_{ij} \left| \sum_{j \in [nx]} y_{ji} - \sum_{j \in [nx]} y_{ij} \geq \zeta_i - x_i, \ i \in [nx] \right. \right\},
$$

where $y \in \mathbb{R}^{nx \times nx^2}$. The uncertain demand $ζ$ is assumed to reside in a budgeted uncertainty set

$$
U = \left\{ \zeta \in \mathbb{R}^{nx^2} \mid \zeta \leq 20, \ 1^\top \zeta \leq 20\sqrt{nx} \right\}.
$$

We pick the $nx$ store locations uniformly at random from $[0, 10]^2$. Two versions of the lot-sizing problem are considered, where the problem with uncapacitated transportation (17) has relatively complete recourse (Section 5.1), while the one with capacitated transportation does not (Section 5.2).

![Figure 1](image.png)

**Figure 1.** A solution the uncapacitated lot-sizing problem with 40 stores. The filled dots have stock and the larger the dots are, the more stock is allocated.
5.1 Uncapacitated transportation

Let the unit cost $t_{ij}$ to transport demand from location $i$ to $j$ be the Euclidean distance if $i \neq j$, and $t_{ii} = 0$, for all $i, j \in [n_x]$. The storage cost per unit is $c_i = 10$, $i \in [n_x]$, and the capacity of each store is 20, i.e., $\mathcal{X} = \{ x \in \mathbb{R}^{n_x} \mid x \leq 20 \}$. All the linear optimization models are solved via MOSEK v9.0 (2019) and run on an Intel XEON CPU with 3.50GHz and 16GB of RAM.

We observe from Table 1 and Table 2 that Algorithm 3 is efficient and effective, i.e., the obtained solutions reduces the average optimality gap of the best known method Bertsimas and de Ruiter (2016) from 3% to 0.1% in a few seconds. Moreover, using the solutions from Algorithm 3 as a warm start for the CCG method reduces the solution time on average because $\text{Time}_d$ is consistently higher than the corresponding $\text{Time}$ in Table 1.

Since the binary variables in the mixed integer linear programming subproblems for the dual formulation, i.e., $2n_x + 1$, is significantly less than that for the primal formulation (17), i.e., $n_x^2 + n_x$, applying the CCG method to (17) leads to a much higher solution time on average than to its dual formulation, i.e., $\text{Time}_p$ is much higher than $\text{Time}_d$ in Table 1.

Ardestani-Jaafari and Delage (2016); Xu and Burer (2018); Hanasusanto and Kuhn (2018) propose to use SDP approximation to conservatively approximate the copositive reformulation of two-stage robust linear optimization models. In Xu and Burer (2018), for the considered uncapacitated lot-sizing problem in Section 5.1, their approach gives solutions of the same quality as the ones from using affine polices. In Georghiou et al. (2017), extensive numerical experiments show that for two-stage robust linear models with relatively complete recourse, the exact solution method of Zeng and Zhao (2013) outperforms the existing methods of Hadjiyiannis et al. (2011); Kuhn et al. (2011); Zhen et al. (2018a); Georghiou et al. (2017). In this paper, we therefore decided to focus on improving applicability and efficiency of the state-of-the-art method of Zeng and Zhao (2013).

5.2 Capacitated transportation

Bertsimas and Shtern (2018) consider a variant version of the lot-sizing problem (17), where the location of each store is randomly picked from a standard 2D Gaussian distribution, the storage cost per unit is $c_i = 1$, $i \in [n_x]$, the transport stock $y_{ij}$ is upper bounded by $20u_{ij}/(N-1)$, and $u_{ij}$ is randomly picked from a standard uniform distribution, $i, j \in [n_x]$, ceteris paribus as in Section 5.1.
Table 1. The average optimality gaps (in %) for the uncapacitated lot-sizing problem with $n_x \in \{3, 5, ..., 15\}$. OBJ-O denotes the optimal value obtained from solving the dual formulation of (17) via the CCG method (given $\hat{x}$ from Algorithm 3 as an initial solution). LB denotes the optimality gaps of the solutions obtained from Algorithm 3, while LB-B and UB are the optimality gaps of using the approach of Bertsimas and de Ruiter (2016) and affine policies, respectively. Time is reported total solution time (in second) for solving the corresponding model, while $\text{Time}_d$ and $\text{Time}_p$ denote the solution times for solving the dual and primal formulations of (17) via the CCG method with an initial solution from (P-S) with scenarios from solving (P) with affine polices. All the numbers are averages of 10 randomly generated instances. “-” in means that the total solution time exceeds the prescribed limit of 2 hours.

<table>
<thead>
<tr>
<th>N</th>
<th>OBJ-O</th>
<th>Time(s)</th>
<th>Time$_d$(s)</th>
<th>Time$_p$(s)</th>
<th>LB</th>
<th>Time(s)</th>
<th>LB-B</th>
<th>Time(s)</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>410</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.1</td>
<td>1.1</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>565</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>5.9</td>
<td>5.9</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>709</td>
<td>0.1</td>
<td>0.4</td>
<td>0.7</td>
<td>0.7</td>
<td>16.2</td>
<td>16.2</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>824</td>
<td>0.2</td>
<td>0.4</td>
<td>1.1</td>
<td>1.1</td>
<td>27.1</td>
<td>27.1</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>930</td>
<td>0.4</td>
<td>0.7</td>
<td>3.6</td>
<td>3.6</td>
<td>54.1</td>
<td>54.1</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1010</td>
<td>0.5</td>
<td>1.1</td>
<td>61.9</td>
<td>61.9</td>
<td>64.1</td>
<td>64.1</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1104</td>
<td>2.0</td>
<td>2.0</td>
<td>4191.7</td>
<td>-</td>
<td>2.4</td>
<td>2.4</td>
<td>5.2</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. The average optimality gaps (in %) for the uncapacitated lot-sizing problem with $n_x \in \{20, \ldots, 45\}$. OBJ-O denotes the optimal value obtained from solving the dual formulation of (17) via the CCG method (given $\hat{x}$ from Algorithm 3 as an initial solution). LB denotes the optimality gaps of the solutions obtained from Algorithm 3, while LB-B and UB are the optimality gaps of using the approach of Bertsimas and de Ruiter (2016) and affine policies, respectively. Time is reported total solution time (in second) for solving the corresponding model. All the numbers are averages of 10 randomly generated instances.

<table>
<thead>
<tr>
<th>N</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJ-O</td>
<td>1279</td>
<td>1445</td>
<td>1576</td>
<td>1747</td>
<td>1857</td>
<td>1979</td>
</tr>
<tr>
<td>Time(s)</td>
<td>25</td>
<td>109</td>
<td>240</td>
<td>965</td>
<td>5413</td>
<td>28293</td>
</tr>
<tr>
<td>LB</td>
<td><strong>0.1</strong></td>
<td><strong>0.2</strong></td>
<td><strong>0.1</strong></td>
<td><strong>0.1</strong></td>
<td><strong>0.0</strong></td>
<td><strong>0.0</strong></td>
</tr>
<tr>
<td>Time(s)</td>
<td>2</td>
<td>1 (54)</td>
<td>5</td>
<td>4</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>LB-B</td>
<td>2.8</td>
<td>2.7</td>
<td>2.1</td>
<td>3.8</td>
<td>4.1</td>
<td>3.9</td>
</tr>
<tr>
<td>Time(s)</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>UB</td>
<td>5.6</td>
<td>5.2</td>
<td>5.6</td>
<td>4.7</td>
<td>4.9</td>
<td>4.6</td>
</tr>
<tr>
<td>Time(s)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
The primal formulation of this capacitated lot-sizing problem has no relatively complete recourse, and thus the CCG method of Zeng and Zhao (2013) cannot be applied to the primal formulation. However, the corresponding dual formulation has complete recourse. Therefore, we propose to solve the dual formulation via the CCG method.

The solutions from enhanced Algorithm 3 produce an average optimality gap of less than 1%, while that of the solutions from Algorithm 3 is around 2%. Note that it takes around 0.1 seconds for the CCG method to reach optimality when $N = 5$ using the dual formulation, while the average solution time is more than 7 seconds for solving the primal formulation using the method of Bertsimas and Shtern (2018). The propose approach is simple to implement while its numerical performance is competitive to the existing methods. For future research, it may be interesting to perform more extensive numerical experiments to compare the proposed approach with recent method of Bertsimas and Shtern (2018).
Table 3. The average optimality gaps (in %) for the capacitated lot-sizing problem with \( n_x \in \{3, 5, \ldots, 15\} \). OBJ-O denotes the optimal value obtained from solving the dual formulation via the CCG method (given \( \hat{x} \) from Algorithm 3 as an initial solution). LB+ and LB denote the optimality gaps of the solutions obtained from Algorithm 3, where LB+ considers additional scenarios from solving (P) with affine policies. LB-B and UB are the optimality gaps of using the approach of Bertsimas and de Ruiter (2016) and affine policies, respectively. Time is reported total solution time (in second) for solving the corresponding model. All the numbers are averages of 10 randomly generated instances.

<table>
<thead>
<tr>
<th>N</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>13</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJ-O</td>
<td>54</td>
<td>91</td>
<td>118</td>
<td>159</td>
<td>185</td>
<td>212</td>
<td>244</td>
</tr>
<tr>
<td>Time(s)</td>
<td>0</td>
<td>0.1</td>
<td>0.4</td>
<td>0.7</td>
<td>6.3</td>
<td>16.7</td>
<td>29.7</td>
</tr>
<tr>
<td>LB+</td>
<td>0.5</td>
<td>1.0</td>
<td>0.7</td>
<td>0.5</td>
<td>0.4</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>Time(s)</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>LB</td>
<td>0.6</td>
<td>1.7</td>
<td>1.4</td>
<td>1.3</td>
<td>1.9</td>
<td>3.4</td>
<td>2.0</td>
</tr>
<tr>
<td>Time(s)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>LB-B</td>
<td>7.6</td>
<td>6.6</td>
<td>3.2</td>
<td>2.3</td>
<td>0.7</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Time(s)</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.6</td>
<td>0.9</td>
<td>1.5</td>
</tr>
<tr>
<td>UB</td>
<td>7.6</td>
<td>6.9</td>
<td>11.8</td>
<td>6.2</td>
<td>10.1</td>
<td>11.5</td>
<td>9.0</td>
</tr>
</tbody>
</table>
References


A Proof for the equivalence of (P') and (D')

\[
\begin{align*}
\min_{x \in \mathcal{X}} \max_{\zeta \in \mathcal{U}'} \min_{y \geq 0} \{ c^\top x + b^\top y \mid A(x)\zeta + By \geq h(x) \} \\
= \min_{x \in \mathcal{X}} \max_{\zeta \in \mathcal{U}'} \max_{w \in \mathcal{W}'} \{ c^\top x + w^\top h(x) - w^\top A(x)\zeta \} \\
= \min_{x \in \mathcal{X}} \max_{w \in \mathcal{W}'} \max_{\zeta \in \mathcal{U}'} \{ c^\top x + w^\top h(x) - w^\top A(x)\zeta \} \\
= \min_{x \in \mathcal{X}} \max_{w \in \mathcal{W}'} \min_{\lambda \geq 0} \{ c^\top x + w^\top h(x) + d^\top \lambda \mid D^\top \lambda \geq -A^\top(x)w \},
\end{align*}
\]

Here the first and the third equality follows from strong duality, which applies because both sets \( \mathcal{U}' \) and \( \mathcal{W}' \) are nonempty by assumption.

B Dual of Fourier-Motzkin elimination: algorithm

Algorithm 1 is adapted from Dantzig and Eaves (1973).

C Progressive solution algorithms
Algorithm 1 Lifting a polyhedral set via the dual of Fourier-Motzkin elimination

Input: \( \mathcal{Z} = \{ z \in \mathbb{R}^n_+ \mid p_i^T z = q_i, \, i \in [m] \} \), where \( p_i \in \mathbb{R}^n \) and \( q_i \in \mathbb{R} \) for every \( i \in [m] \).

**Step 1.** Create an equation with a zero right-hand-side

\[
\sum_{j \in J^+} \hat{p}_{1j} z_j - \sum_{j \in J^-} \hat{p}_{1j} z_j = 0,
\]

where \( \hat{p}_{1j} = p_{1j}, \, j \in J^+, \) \( \hat{p}_{1j} = -p_{1j}, \, j \in J^- \), i.e., \( \hat{p}_{1j} \geq 0, \, j \in [n_z], \) \( J^+ \cap J^- = \emptyset \), and \( J^+, J^- \subseteq [n_z] \). This equation can be obtained by combining the 1-st constraint with some appropriate linear combination of the equalities in \( \mathcal{Z} \). Then, we find

\[
\mathcal{Z} = \left\{ z \in \mathbb{R}^n_+ \mid \sum_{j \in J^+} \hat{p}_{1j} z_j - \sum_{j \in J^-} \hat{p}_{1j} z_j = 0, \, p_i^T z = q_i, \, i \in [m] \setminus \{1\} \right\}.
\]

**Step 2.** Introduce \(|J^+| \times |J^-|\) new variables \( w_{jk} \geq 0 \) by setting

\[
z_j = \begin{cases} \sum_{k \in |J^-|} w_{jk}/\hat{p}_{1j} & j \in J^+ \\ \sum_{k \in |J^+|} w_{kj}/\hat{p}_{1j} & j \in J^- \end{cases}.
\]

The set \( \mathcal{Z} \) can now be lifted as follows

\[
\mathcal{Z}' = \left\{ z \in \mathbb{R}^n_+ \mid \begin{array}{l}
p_i^T z = q_i, \, i \in [m] \setminus \{1\} \\
z_j = \sum_{k \in |J^-|} w_{jk}/\hat{p}_{1j}, \, j \in J^+ \\
z_j = \sum_{k \in |J^+|} w_{kj}/\hat{p}_{1j}, \, j \in J^- \end{array} \right\}.
\]

Output: \( \mathcal{Z}' \)
**Algorithm 2 Mountain climbing v.1**

**Input:** \( \{\zeta^{(1)}, \ldots, \zeta^{(|\mathcal{U}|)}\} \leftarrow \mathcal{U} \subset \mathcal{U} \)

Candidate solution \( \hat{x} \)

Tolerance \( \varepsilon > 0 \)

Set \( \Delta \leftarrow 1, \mathcal{W} \leftarrow \emptyset \) (Initialization)

repeat

Set \( k \leftarrow 1 \)

repeat

\( \hat{w} \leftarrow \text{Solve (15) given } (\hat{x}, \zeta^{(k)}) \)

if \( v_{\hat{w}} > 0 \) given \( (\hat{x}, \hat{\zeta}) \) then

Set \( \mathcal{W} \leftarrow \hat{w} \cup \mathcal{W} \)

end if

\( \hat{\zeta} \leftarrow \text{Solve (16) given } (\hat{x}, \hat{w}) \)

if \( v_{\hat{\zeta}} > 0 \) given \( (\hat{x}, \hat{w}) \) then

Set \( \mathcal{U} \leftarrow \hat{\zeta} \cup \mathcal{U} \)

end if

Set \( k \leftarrow k + 1 \)

until \( k = |\mathcal{U}| \)

Set \( \hat{\mathcal{U}} \leftarrow c^T \hat{x} \)

\( \hat{x} \leftarrow \text{Solve (PD-S) given } \overline{\mathcal{U}} \text{ and } \overline{\mathcal{W}} \)

Set \( \Delta \leftarrow (c^T \hat{x} - LB)/LB \)

until \( \Delta < \varepsilon \)

Output: \( LB \)

---

**Algorithm 3 Mountain climbing v.2**

**Input:** \( \{w^{(1)}, \ldots, w^{(|\mathcal{W}|)}\} \leftarrow \mathcal{W} \subset \mathcal{W} \)

Candidate solution \( \hat{x} \)

Tolerance \( \varepsilon > 0 \)

Set \( \Delta \leftarrow 1, \mathcal{W} \leftarrow \emptyset, \mathcal{U} \leftarrow \emptyset \) (Initialization)

repeat

Set \( k \leftarrow 1 \)

repeat

\( \hat{\zeta} \leftarrow \text{Solve (16) given } (\hat{x}, w^{(k)}) \)

if \( v_{\hat{\zeta}} > 0 \) then

\( \hat{w} \leftarrow \text{Solve (15) given } (\hat{x}, \hat{\zeta}) \)

Set \( \hat{\mathcal{W}} \leftarrow \hat{w} \cup \hat{\mathcal{W}}, \hat{\mathcal{U}} \leftarrow \hat{\zeta} \cup \hat{\mathcal{U}} \)

end if

Set \( k \leftarrow k + 1 \)

until \( k = |\mathcal{W}| \)

Set \( LB \leftarrow c^T \hat{x}, \overline{\mathcal{W}} \leftarrow \hat{\mathcal{W}} \)

\( \hat{x} \leftarrow \text{Solve (P-S) given } \hat{\mathcal{U}} \)

Set \( \Delta \leftarrow (c^T \hat{x} - LB)/LB, \overline{\mathcal{W}} \leftarrow \emptyset \)

until \( \Delta < \varepsilon \)

Set \( LB \leftarrow c^T \hat{x} \)

Output: \( LB \)