Duality in Two-stage Adaptive Linear Optimization: Faster Computation and Stronger Bounds

Dimitris Bertsimas
Operations Research Center and Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, dbertsim@mit.edu,

Frans J.C.T. de Ruiter*
Department of Econometrics and Operations Research, Tilburg University, 5000 LE Tilburg, The Netherlands,, f.j.c.t.deruiter@tilburguniversity.edu,

In this paper we derive and exploit duality in general two-stage adaptive linear optimization models. The equivalent dualized formulation we derive is again a two-stage adaptive linear optimization model. Therefore, all existing solution approaches for two-stage adaptive models can be used to solve or approximate the dual formulation. The new dualized model differs from the primal formulation in its dimension and uses a different description of the uncertainty set. We show that the optimal primal affine policy can be directly obtained from the optimal affine policy in the dual formulation. We provide empirical evidence that the dualized model in the context of two-stage lot-sizing on a network and two-stage facility location problems solves an order of magnitude faster than the primal formulation with affine policies. We also provide an explanation and associated empirical evidence that offer insight on which characteristics of the dualized formulation make computations faster. Furthermore, the affine policy of the dual formulations can be used to provide stronger lower bounds on the optimality of affine policies.

Key words: two-stage problems; robust optimization; duality; affine control policies;

1. Introduction

Many applications for decision making under uncertainty can be naturally modeled as two-stage adaptive optimization models. In these models some of the decisions have to be made *here-and-now* before the realization of the uncertain parameter is known. The other decisions are of a *wait-and-see* type, which are chosen after the realization of the uncertain parameter is known. One way of dealing with these problems is via stochastic optimization. These methods assume that a probabilistic description of the realization is known and optimize for expected values. For references on these techniques we refer to Birge and Louveaux (2011) and Kali and Wallace (1994). Stochastic models, especially in a

* This research was conducted while the second author was visiting the Operations Research Center at the Massachusetts Institute of Technology and was partly financed by the Netherlands Organisation for Scientific Research (NWO) Research Talent grant 406-14-067.

two-stage setting, are known to suffer from the 'curse of dimensionality' and are therefore likely not tractable, see e.g. Shapiro and Nemirovski (2005). A different approach is to model these two-stage problems in a robust setting. Robust optimization techniques do not require a probabilistic description of the uncertainty set and have proven to be very useful in a number of practical applications. A selection of applications that use a two-stage robust setting are: unit commitment in the energy sector (Bertsimas et al. 2013, Wang et al. 2013, Zhao and Zeng 2012), emergency supply chain planning (Ben-Tal et al. 2011), facility location problems (Ardestani-Jaafari and Delage 2014, Atamtürk and Zhang 2007, Gabrel et al. 2014a), Capacity expansion of network flows (Ordóñez and Zhao 2007, Yin et al. 2009) and many others, see e.g. the survey papers by Bertsimas et al. (2011) and Gabrel et al. (2014b).

In the last decade or so, there has been a rise in solution techniques tailored to solve two-stage optimization models in a robust setting. One of the first and very popular method is the use of affine policies for the wait-and-see decisions proposed by Ben-Tal et al. (2004). This method is appealing because it is computationally tractable for problem instances of moderate to large size. Furthermore, the affine policies appear to be near optimal in practical applications (Ardestani-Jaafari and Delage 2014, Ben-Tal et al. 2004, 2005). The use of affine policies is even provably optimal in some special cases (Bertsimas et al. 2010, Iancu et al. 2013). Other methods designed to solve two-stage adaptive optimization models are: approximation by static solutions (Bertsimas and Goyal 2010), finite adaptability (Bertsimas and Caramanis 2010), enumeration of vertices of the uncertainty set (Bertsimas and Goyal 2012), column generation algorithms (Zeng and Zhao 2013) and iterative partitioning of the uncertainty set (Postek and Den Hertog 2014, Bertsimas and Dunning 2014).

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

1. We provide a dualized two-stage two-stage adaptive model for linear two-stage models with continuous wait-and-see decisions. The new model is derived by consecutively dualizing over the wait-and-see decisions and the uncertain parameters. The new dualized formulations have the same set of feasible (and optimal) here-and-now decisions as the original two-stage models. It has different dimensions, uncertain parameters, wait-and-see

decisions and constraints than the original two-stage adaptive model. Since the model is again a two-stage adaptive model, all existing solution techniques for two-stage adaptive models can be used to solve it.

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

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

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

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

The rest of this paper is organized as follows. In Section 2, we introduce the two-stage adaptive optimization model and derive the new dualized two-stage model. We explain the use of affine policies in the primal and dual formulation in Section 3. Section 4 gives the computational algorithm to obtain stronger bounds on the optimal value of the fully adaptive model. In Sections 5 and 6, we present our numerical results and show the computational advantage of the dualized formulation. Section 7 gives some concluding remarks.

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

2. Duality in Two-stage Adaptive Formulations

We first state the usual two-stage formulation in Section 2.1. The new dualized formulation is given in Section 2.2. We also indicate similarities in structure with the primal formulation and the differences in the two formulations.

2.1. The Primal Formulation

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

$$\mathcal{U} = \{\boldsymbol{\zeta} \geq \boldsymbol{0} : \ \boldsymbol{D}\boldsymbol{\zeta} \leq \boldsymbol{d}\}, \tag{1}$$

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

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U} : \ \exists \boldsymbol{y} \geq \boldsymbol{0} : \ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2}$$

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

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

for some matrices $\boldsymbol{R}_0, \boldsymbol{R}_1, \dots, \boldsymbol{R}_n$. For our dual derivation to work, we must have the matrix $\boldsymbol{B}$ to be fixed independent of $\boldsymbol{\zeta}$. Hence, we only consider the case of *fixed recourse*. Without loss of generality, there is no uncertainty in the objective function and it only includes here-and-now decisions. Objectives including uncertain parameters and wait-and-see decisions can be modelled as an instance of (2) using an epigraph formulation, see (Ben-Tal et al. 2009, pp. 10-11). These epigraph formulations are also used in the models of our numerical examples in Sections 5 and 6.

2.2. The New Dualized Formulation

The main contributions of this paper come from the next theorem, giving a dual formulation of (2).

THEOREM 1. *The here-and-now decision $\boldsymbol{x}$ is feasible (and optimal) for* (2) *with nonempty uncertainty set $\mathcal{U}$ as in* (1) *if and only if $\boldsymbol{x}$ is feasible (and optimal) for*

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{w} \in \mathcal{V}: \ \exists \boldsymbol{\lambda} \geq \boldsymbol{0}: \begin{cases} \boldsymbol{w}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{r}) - \boldsymbol{d}^\top \boldsymbol{\lambda} \geq 0 \\ \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \boldsymbol{R}^\top \boldsymbol{w} \end{cases} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{3}$$

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

The proof of this theorem is split in two parts. The first part comes from a result known in the literature and the second part is the new contribution leading to the dualized formulation. The result from the literature transforms (2) into a bilinear optimization model by applying duality to the wait-and-see variables. The result from this part is used frequently in the literature, in various settings, to solve two-stage adaptive optimization problems using column generation and Benders decomposition type algorithms (see e.g. Bertsimas et al. (2013), Minoux (2011), Thiele et al. (2009), Zeng and Zhao (2013) and Zhao and Zeng (2012)) or to derive an exact solution for special cases (Ordóñez and Zhao 2007). This known result is given in Lemma 1.

LEMMA 1. *The here-and-now decision $\boldsymbol{x}$ is feasible (and optimal) for* (2) *if and only if $\boldsymbol{x}$ is feasible (and optimal) for*

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\zeta} \in \mathcal{U}} \max_{\boldsymbol{w} \geq \boldsymbol{0}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{w}^\top (\boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) \mid \boldsymbol{B}^\top \boldsymbol{w} \leq \boldsymbol{0} \right\}. \tag{4}$$

Proof. For a given $\boldsymbol{x} \in \mathcal{X}$ we can write (2) as

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

The result then follows by dualizing over $\boldsymbol{y}$. $\square$

Note that for every $\boldsymbol{\zeta}$ the variable $\boldsymbol{w}$ ensures that the problem returns ∞ whenever there exists a $\boldsymbol{\zeta}$ that violates the constraints in the original model (2). The result from Lemma 1 is also used in Kuhn et al. (2011) to assess the suboptimality of affine policies in a two-stage *stochastic* setting. Their bound can also be used in robust settings, but one has to assign a distribution to the uncertainty set a priori. The authors explain that in that case the quality of the bound depends on the a priori distribution that is chosen. For the rest of the proof we first dualize (4) further to end up with an equivalent two-stage adaptive optimization formulation.

Proof of Theorem 1. Consider, for fixed $\boldsymbol{w}$, the inner maximization problem in (4). Dualizing over $\boldsymbol{\zeta}$ gives

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{w}\geq\boldsymbol{0}} \min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^\top\boldsymbol{x} + \boldsymbol{w}^\top(\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top\boldsymbol{\lambda} \mid \boldsymbol{D}^\top\boldsymbol{\lambda} \geq \boldsymbol{R}^\top\boldsymbol{w},\ \boldsymbol{B}^\top\boldsymbol{w} \leq \boldsymbol{0} \right\} \\ &\quad = \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{w}\in\widetilde{\mathcal{V}}} \min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^\top\boldsymbol{x} + \boldsymbol{w}^\top(\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top\boldsymbol{\lambda} \mid \boldsymbol{D}^\top\boldsymbol{\lambda} \geq \boldsymbol{R}^\top\boldsymbol{w} \right\}, \end{aligned} \tag{5}$$

where in the last line we introduced $\widetilde{\mathcal{V}} = \left\{\boldsymbol{w} \geq \boldsymbol{0} :\ \boldsymbol{B}^\top\boldsymbol{w} \leq \boldsymbol{0}\right\}$. Introducing a variable γ we write the model using an epigraph formulation

$$\begin{aligned} \min_{\boldsymbol{x},\gamma} \quad & \boldsymbol{c}^\top\boldsymbol{x} + \gamma \\ \text{s.t.} \quad & \forall \boldsymbol{w} \in \widetilde{\mathcal{V}} :\ \exists \boldsymbol{\lambda} \geq \boldsymbol{0} : \begin{cases} \boldsymbol{w}^\top(\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top\boldsymbol{\lambda} \leq \gamma \\ \boldsymbol{D}^\top\boldsymbol{\lambda} \geq \boldsymbol{R}^\top\boldsymbol{w} \end{cases} \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{6}$$

To end up with our final result (3) we have to prove that $\gamma = 0$ for any optimal solution and that we can add the additional restriction $\boldsymbol{e}^\top\boldsymbol{w} = 1$ to bound the uncertainty set $\widetilde{\mathcal{V}}$ without affecting the set of feasible solutions. From (5) it follows that there has to be an optimal adaptive policy $\boldsymbol{\lambda}^*(\boldsymbol{w})$ that satisfies

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

Note that this policy is not only worst-case optimal, but chooses the best wait-and-see decision $\boldsymbol{\lambda}(\boldsymbol{w})$ for every scenario $\boldsymbol{w}$. For the scenario $\mathbf{0} \in \widetilde{\mathcal{V}}$ we have

$$\boldsymbol{d}^\top (\boldsymbol{\lambda}^*(\mathbf{0})) = \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \{\boldsymbol{d}^\top \boldsymbol{\lambda} \mid \boldsymbol{D}^\top \boldsymbol{\lambda} \geq \mathbf{0}\} = \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \max_{\boldsymbol{\zeta} \geq \mathbf{0}} \{\boldsymbol{\lambda}^\top (\boldsymbol{d} - \boldsymbol{D}\boldsymbol{\zeta})\} = 0,$$

where the last equality holds since $\mathcal{U} = \{\boldsymbol{\zeta} \geq \mathbf{0} : \ \boldsymbol{D}\boldsymbol{\zeta} \leq \boldsymbol{d}\}$ is nonempty. Using this optimal decision for the parameter $\mathbf{0} \in \widetilde{\mathcal{V}}$, we see that

$$\gamma \geq \mathbf{0}^\top (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top \boldsymbol{\lambda}^*(\mathbf{0}) = 0. \tag{7}$$

Now, let $t \geq 0$ and $\boldsymbol{w} \geq \mathbf{0}$. Then we have

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

Hence, we can impose scalar multiplicity on the adaptive policy $\boldsymbol{\lambda}^*(\boldsymbol{w})$ without affecting the value of $\boldsymbol{d}^\top (\boldsymbol{\lambda}^*(\boldsymbol{w}))$. That is, for every $\boldsymbol{w} \in \widetilde{\mathcal{V}}$ and scalar $t \geq 0$ we impose $\boldsymbol{\lambda}^*(t\boldsymbol{w}) = t\boldsymbol{\lambda}(\boldsymbol{w})$. From (7) we have that $\gamma \geq 0$. Suppose for the sake of contradiction that for an optimal here-and-now decision $(\boldsymbol{x}, \gamma)$ we have $\gamma > 0$. Then there exist an $\boldsymbol{w} \in \widetilde{\mathcal{V}}$, $\boldsymbol{w} \neq \mathbf{0}$ such that

$$\boldsymbol{w}^\top (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top \boldsymbol{\lambda}^*(\boldsymbol{w}) = \gamma > 0.$$

Since $\widetilde{\mathcal{V}}$ is a cone, we have that $(t\boldsymbol{w}) \in \widetilde{\mathcal{V}}$ for every $t \geq 0$ and $\boldsymbol{w} \in \widetilde{\mathcal{V}}$. Therefore, we have by scalar multiplicity of $\boldsymbol{\lambda}^*(\boldsymbol{w})$

$$(t\boldsymbol{w})^\top (\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}) + \boldsymbol{d}^\top \boldsymbol{\lambda}^*(t\boldsymbol{w}) = t\gamma > \gamma \quad \text{for all} \quad t > 1.$$

This contradicts the assumption that $\gamma > 0$ is feasible. Hence, we must have $\gamma = 0$. Finally, consider a solution that is feasible for all values in the further restricted uncertainty set

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

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

Any two-stage adaptive optimization model with fixed recourse, continuous wait-and-see decisions and a polyhedral uncertainty set can be readily formulated as an instance of (2). Theorem 1 then directly provides practitioners with the alternative dual formulation (3). Table 1 highlights some differences such as the number of wait-and-see variables, uncertain parameters and constraints in the primal and dual formulation. In our numerical examples in Sections 5 and 6 we clarify these differences with explicit values for m, k, L and p.

Table 1 **Comparing dimensions of uncertainty parameters, variables and number of constraints in the original two-stage adaptive formulation** (2) **and in our new dualized formulation** (3)**.**

	Primal formulation (2)	Dual formulation (3)
# uncertain parameters	L	m
# wait-and-see decisions	k	p
# constraints on variables	m	$L+1$
# constraints on uncertain parameter	p	$k+1$

3. Solving the Primal and Dual Formulation with Affine Policies

The model (3) is again a two-stage adaptive optimization model with a nonnegative bounded polyhedral uncertainty set and is therefore another instance of (2). Hence, we can directly apply all exact and approximation methods to solve adaptive optimization problems mentioned in the introduction. We first show the equivalence of the dual formulation with the nonadaptive robust counterpart in the static case. We then continue to show that the optimal solutions of both formulations are the same when we solve the models with affine policies.

3.1. Static Robust Optimization

If we take $\boldsymbol{B}=\boldsymbol{O}$, then (2) is the following robust optimization model without wait-and-see decisions:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U}: \; \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{8}$$

where $\mathcal{U}$ is as in (1). This problem is hard to solve in its current form since each constraint has to hold for an infinite number of values for $\boldsymbol{\zeta}$. To reformulate the problem, we can consider the uncertainty constraintwise (see Ben-Tal et al. (2009)), i.e., we only have to look at one row

$$\forall \boldsymbol{\zeta} \in \mathcal{U}: \; \boldsymbol{A}_i \boldsymbol{x} \geq \boldsymbol{R}_i \boldsymbol{\zeta} + r_i \tag{9}$$

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

Lemma 2 **(Robust Counterpart)**. *Constraint* (9) *is satisfied if and only if there exists a* $\boldsymbol{\pi}^{\boldsymbol{i}} \in \mathbb{R}^p$ *such that*

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

Note that this dualization approach can also be used for any other polyhedral uncertainty set. For notational convenience we shall use matrix variables for the rest of the section. If we write $\boldsymbol{\Pi} = [\boldsymbol{\pi}^{\boldsymbol{i}}, \ldots, \boldsymbol{\pi}^{\boldsymbol{m}}]$, then by Lemma 2 we have that (8) is equivalent to

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{\Pi}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{\Pi}^\top \boldsymbol{d} \geq \boldsymbol{r} \\
& \boldsymbol{D}^\top \boldsymbol{\Pi} \geq \boldsymbol{R}^\top \\
& \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\Pi} \geq \boldsymbol{O}.
\end{aligned} \tag{10}$$

We can also find a dual formulation for the static model (8) using the dual formulation that is derived in Theorem 1. In that way, we end up with the same dual formulation as in (3), but with the simple uncertainty set

$$\mathcal{V} = \left\{ \boldsymbol{w} \geq \boldsymbol{0} : \ \boldsymbol{e}^\top \boldsymbol{w} = 1 \right\}. \tag{11}$$

For these robust models with $\boldsymbol{B} = \boldsymbol{O}$ the uncertainty set (11) has only m extreme points $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_m}$. As shown in (Bertsimas and Goyal 2012, Lemma 1), linear policies are optimal if there are only m extreme points, where m is the size of the uncertainty set. Furthermore, by taking the linear policy $\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{\Pi}\boldsymbol{w}$ in (3) we end up with the same robust counterpart as (10).

3.2. Solving the Two-stage Formulations with Affine Policies

Let us now return to the general case in which $\boldsymbol{B} \neq \boldsymbol{O}$, so we do need to take the wait-and-see decisions $\boldsymbol{y}$ into account. In principle, an optimal policy $\boldsymbol{y}(\boldsymbol{\zeta})$ in (2) can be any function of the uncertain parameter $\boldsymbol{\zeta}$. However, this results in an intractable model where we would have to optimize over all possible functions. To come up with tractable models Ben-Tal et al. (2004) suggest to restrict the wait-and-see decisions to be affine in $\boldsymbol{\zeta}$:

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

where $\boldsymbol{u} \in \mathbb{R}^m$ and $\boldsymbol{V} \in \mathbb{R}^{m\times L}$ are respectively a vector and a matrix of here-and-now variables. Although this restriction might seem very severe, it turns out to perform very good in practical applications, see Ben-Tal et al. (2004, 2005), and is even provably optimal in some specific cases, see Bertsimas et al. (2010), Iancu et al. (2013). With this decision rule, we obtain the following robust counterpart for (2) with affine policies

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{V}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U}: \begin{cases} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{u}+\boldsymbol{V}\boldsymbol{\zeta}) \geq \boldsymbol{R}\boldsymbol{\zeta} + \boldsymbol{r} \\ \boldsymbol{u} + \boldsymbol{V}\boldsymbol{\zeta} \geq \boldsymbol{0} \end{cases} \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{12}$$

This model does not have wait-and-see variables. Therefore, we can apply Lemma 2 to reformulate each constraint and obtain the robust counterpart. Introducing the auxiliary (matrix) variables $\boldsymbol{\Pi} \in \mathbb{R}^{p\times m}$ and $\boldsymbol{\Xi} \in \mathbb{R}^{p\times k}$ we can write down the robust counterpart as

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{u},\boldsymbol{V},\boldsymbol{\Pi},\boldsymbol{\Xi}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} - \boldsymbol{\Pi}^\top \boldsymbol{d} \geq \boldsymbol{r} \\ & \boldsymbol{B}\boldsymbol{V} \geq \boldsymbol{R} - \boldsymbol{\Pi}^\top \boldsymbol{D} \\ & \boldsymbol{u} - \boldsymbol{\Xi}^\top \boldsymbol{d} \geq \boldsymbol{0} \\ & \boldsymbol{D}^\top \boldsymbol{\Xi} + \boldsymbol{V}^\top \geq \boldsymbol{O} \\ & \boldsymbol{\Pi}, \boldsymbol{\Xi} \geq \boldsymbol{O} \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{13}$$

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

$$\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{Q}\boldsymbol{w}, \tag{14}$$

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

Proposition 1. *If* $(\boldsymbol{x},\ \boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{q} + \boldsymbol{Q}\boldsymbol{w})$ *is feasible for* (3), *then* $(\boldsymbol{x}, \widetilde{\boldsymbol{\lambda}}(\boldsymbol{w}) = \widetilde{\boldsymbol{Q}}\boldsymbol{w})$ *with* $\widetilde{\boldsymbol{Q}} = \boldsymbol{q}\boldsymbol{e}^\top + \boldsymbol{Q}$ *is also feasible.*

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

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

Hence, if $\boldsymbol{\lambda}(\boldsymbol{w})$ is a feasible policy for (15), then so is $\widetilde{\boldsymbol{\lambda}}(\boldsymbol{w})$. $\square$

Substituting the linear policy (14) in (3), we obtain the following model

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{Q}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \forall \boldsymbol{w} \in \mathcal{V}: \begin{cases} \boldsymbol{w}^\top(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{r}) - \boldsymbol{d}^\top(\boldsymbol{Q}\boldsymbol{w}) \geq \boldsymbol{0} \\ \boldsymbol{D}^\top \boldsymbol{Q}\boldsymbol{w} \geq \boldsymbol{R}^\top \boldsymbol{w} \\ \boldsymbol{Q}\boldsymbol{w} \geq \boldsymbol{0} \end{cases} \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{15}$$

A robust counterpart for (15) can be derived using standard LP dualization as in Lemma 2. With the introduction of the auxiliary variables $\boldsymbol{\varepsilon} \in \mathbb{R}^k$, $\boldsymbol{\Lambda} \in \mathbb{R}^{k\times L}$ and $\boldsymbol{\Omega} \in \mathbb{R}^{k\times p}$, the resulting robust counterpart can be written as

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{Q},\boldsymbol{\varepsilon},\boldsymbol{\Omega},\boldsymbol{\Lambda}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{\varepsilon} - \boldsymbol{Q}^\top \boldsymbol{d} \geq \boldsymbol{r} \\ & \boldsymbol{B}\boldsymbol{\Lambda} \geq \boldsymbol{R} - \boldsymbol{Q}^\top \boldsymbol{D} \\ & \boldsymbol{B}\boldsymbol{\Omega} + \boldsymbol{Q}^\top \geq \boldsymbol{O} \\ & \boldsymbol{\varepsilon} \geq \boldsymbol{0},\ \boldsymbol{\Lambda}, \boldsymbol{\Omega} \geq \boldsymbol{O} \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{16}$$

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

Theorem 2. *The solution* $(\boldsymbol{x}, \boldsymbol{Q}, \boldsymbol{\varepsilon}, \boldsymbol{\Omega}, \boldsymbol{\Lambda})$ *is feasible for* (16) *if and only if* $(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{V}, \boldsymbol{\Pi}, \boldsymbol{\Xi})$ *is feasible for* (13), *where*

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

$$\mathbf{\Xi} = \mathbf{\Omega}^\top.$$

The proof is direct and therefore omitted. Theorem 2 is not only useful because it proves equivalence of the primal and dual formulation with affine policies. It also allows us to solve the dual formulation (16) with affine policies and directly obtain the optimal affine policy of the original formulation (or vice versa). Despite this equivalence there may be significant computational benefits from solving two stage problems using the dualized formulation rather than the primal formulation. This can be seen by comparing the two robust counterparts (13) and (16). We compare the number of affine constraints and the number of sign restrictions in Table 2. We use the same parameters as in Table 1 for the number of uncertain parameters (L), the number of wait-and-see decisions (k), the number of affine constraints on the variables (m) and the number of affine constraints in the uncertainty set (p). We observe that the total number of constraints (affine constraints and sign restrictions) is the same in both formulations. However, there is a difference in the break down into the number of affine constraints and the number of sign restrictions. This is important since sign restrictions are much easier to handle by solvers than affine constraints. From Table 2 we see that for a large number of wait-and-see decisions k, relative to the number of constraints in the original model and uncertainty set (m and p), the dual formulation (16) can most likely be solved more efficiently than the primal formulation (13). We observe these computational benefits in our numerical examples in Sections 5 and 6 where we present Table 2 with some explicit values for L, k, m and p.

Table 2 Comparing the number of affine constraints and sign restrictions in (13) **and** (16)

	Primal formulation (13)	Dual formulation (16)
# affine constraints	$(1+L)(m+k)$	$m(1+L+p)$
# sign restrictions	$p(m+k)$	$k(1+L+p)$

Finally, we note that the models (12) and (15) can also be solved via cutting plane methods, see Mutapcic and Boyd (2009). There have been extensive numerical studies that show that in some cases cutting plane algorithms require slightly less computation time than solving the robust counterpart constructed by Lemma 2 (Fischetti and Monaci 2012, Bertsimas et al. 2015). We have also solved our numerical examples with the cutting plane algorithm described in those papers. As with the reformulation approach, we observe that the dual formulation (15) can be solved an order of magnitude faster than the primal problem. This approach is however not elaborated further for two reasons. First, to construct the primal solution from the dual solution by Theorem 2 we need the auxiliary variables that are introduced by the reformulation. Second, initial findings showed that the cutting plane algorithm is a lot slower for the problems considered Sections 5 and 6. We were only able to solve the smaller instances in reasonable time via cutting planes.

4. Stronger Bounds on the Optimality Gap of Affine policies

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

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

$$\begin{aligned}
\min_{\boldsymbol{x}, \boldsymbol{y}^{\mathbf{1}}, \ldots, \boldsymbol{y}^{\overline{\boldsymbol{N}}}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^{\boldsymbol{i}} \geq \boldsymbol{R}\boldsymbol{\zeta}^{\boldsymbol{i}} + \boldsymbol{r} \qquad \forall i = 1, \ldots, \overline{N} \\
& \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y}^{\mathbf{1}}, \ldots, \boldsymbol{y}^{\overline{\boldsymbol{N}}} \geq 0.
\end{aligned} \tag{17}$$

The crucial question is of course which scenarios to include. It is shown by Bertsimas and Goyal (2012) that the lower bound is tight if we include all extreme points of the uncertainty set. This is in practice undoable since there can be a huge number of extreme points, each resulting in an extra variable and constraint in (17). Another straightforward way would be to sample $\overline{N}$ scenarios uniformly at random from $\mathcal{V}$. The model (17) remains tractable for relatively large $\overline{N}$, but for all our examples we obtain useless bounds, even when the number of random samples $\overline{N}$ is as big as 10^5. We therefore have to pick the scenarios in a more specific way. To do so, we first introduce the notion of *binding scenarios*.

Definition 1 (Binding scenarios). Let $f:\mathcal{U}\times\mathcal{X}\rightarrow\mathbb{R}$ be a function of the uncertain parameter $\boldsymbol{\zeta}\in\mathcal{U}$ and here-and-now decision $\boldsymbol{x}\in\mathcal{X}$. For a given $\boldsymbol{x}\in\mathcal{X}$ the parameter $\widehat{\boldsymbol{\zeta}}$ is called *binding* for the robust constraint

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

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

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

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

$$\begin{array}{lll}\min\limits_{\boldsymbol{x},\boldsymbol{\lambda}^{\boldsymbol{1}},\dots,\boldsymbol{\lambda}^{\overline{\boldsymbol{M}}}} & \boldsymbol{c}^\top\boldsymbol{x} & \\ \text{s.t.} & (\boldsymbol{w}^{\boldsymbol{j}})^\top(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{r})-\boldsymbol{d}^\top\boldsymbol{\lambda}^{\boldsymbol{j}}\geq\boldsymbol{0} \qquad \forall j=1,\dots,\overline{M} & \\ & \boldsymbol{D}^\top\boldsymbol{\lambda}^{\boldsymbol{j}}\geq\boldsymbol{R}^\top\boldsymbol{w}^{\boldsymbol{j}} \qquad \forall j=1,\dots,\overline{M} & \\ & \boldsymbol{x}\in\mathcal{X},\boldsymbol{\lambda}^{\boldsymbol{1}},\dots,\boldsymbol{\lambda}^{\overline{\boldsymbol{M}}}\geq\boldsymbol{0}. & \end{array} \tag{18}$$

Combining the constraints from (17) and (18) results in a model that provides a stronger lower bound than the one that only uses the binding scenarios from the primal formulation. We can now give Algorithm 1 that provides the strengthened bound on the optimal value of the fully adaptive model. Step 1 provides us with a feasible solution and an upper bound on the optimal value of the fully adaptive problem. The objective value of the model in step

Algorithm 1 Stronger bounds on optimality of affine policies

1: Solve (13) to get optimal optimal here-and-now $\boldsymbol{x}$, affine policy $\boldsymbol{y}(\boldsymbol{\zeta}) = \boldsymbol{u} + \boldsymbol{V}\boldsymbol{\zeta}$ and auxiliary variables $\boldsymbol{\Pi}, \boldsymbol{\Xi}$.

2: Construct the dual affine policy $\boldsymbol{\lambda}(\boldsymbol{w}) = \boldsymbol{Q}\boldsymbol{w}$ using Theorem 2.

3: Find the binding scenarios $\{\boldsymbol{\zeta}^{\boldsymbol{1}}, \boldsymbol{\zeta}^{\boldsymbol{2}}, \ldots, \boldsymbol{\zeta}^{\overline{\boldsymbol{N}}}\}$ in (12) and $\{\boldsymbol{w}^{\boldsymbol{1}}, \boldsymbol{w}^{\boldsymbol{2}}, \ldots, \boldsymbol{w}^{\overline{\boldsymbol{M}}}\}$ in (15).

4: Solve the sampled problem with binding scenarios for the primal and dual

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

4 gives us the new lower bound. A binding scenario for each constraints in (12) and (15) can be found directly using the optimal affine policies from step 1 and 2. We omit here the elaborate description of a more efficient way to finding the set of binding scenarios in step 3 via KKT conditions which is described in Hadjiyiannis et al. (2011). However, step 3 is not the most time consuming step as solving the model with affine policies in step 1 takes by far the most time. Finally, we note that we can also solve the dual formulation (16) with affine policies in step 1 and obtain the primal affine policy in step 2 using Theorem 2.

5. Example 1: Lot-sizing on a Network

In this section we present a natural example in which (15) takes an order of magnitude less time to solve than the primal formulation (12). Also, the new lower bound on the fully adaptive model (2) derived from Algorithm 1 is much stronger than the lower bound from Hadjiyiannis et al. (2011) that only used the binding scenarios from the primal formulation.

5.1. Problem Setting

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

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

After we observe the realization of the demand we can transport stock y_{ij} from store i to store j at cost t_{ij} in order to meet all demand. The aim is to minimize the worst case storage costs (with unit costs c_i) and the cost arising from shifting the products from one store to another. This network flow model can now be written as a specific instance of the primal problem (2) as follows:

$$\begin{aligned} \min_{\boldsymbol{x},\alpha} \quad & \alpha \\ \text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U} : \exists \boldsymbol{y} \geq \mathbf{0} : \left\{ \begin{array}{ll} \alpha \geq \sum_{i=1}^N c_i x_i + \sum_{i=1}^N \sum_{j=1}^N t_{ij} y_{ij} & \\ \zeta_i \leq \sum_{j=1}^N y_{ji} - \sum_{j=1}^N y_{ij} + x_i & i = 1, \ldots, N \end{array} \right. \\ & \qquad\qquad\qquad\qquad\qquad\qquad 0 \leq x_i \leq K_i \qquad i = 1, \ldots, N, \end{aligned} \tag{19}$$

where the first line in (19) is for the epigraph formulation. The second line contains the balance equations: we have to shift stock to and from node i such that the initial storage plus the net shift in stock still exceeds the demand at node i. The last constraints restrict the capacity of the stock at each node. Note that this model can be seen as a network flow model with multiple sources and multiple sinks.

5.2. Test Case and Numerical Results

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

We solve both (13) and (15) and depict the average solution times over 10 runs in Table 3, as well as the objective value and the lower bounds. The stock allocation

Table 3 Compare performance of primal and dualized formulation with affine policies for the lot-sizing example. The percentages in the last columns depict the optimality gap derived from each lower bound compared to the objective value. All results are averaged over 10 runs.

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

(the here-and-now decision) for the $N = 30$ instance is depicted in Figure 1. The lower bound from the primal is obtained using the method from Hadjiyiannis et al. (2011). The primal/dual bound is the strengthened bound resulting from Algorithm 1. Solving the model via the new dualized formulation (16) reduces the computation an order of magnitude compared with the original primal formulation (13). For the larger instances we see that the primal formulation is approximately 20 times slower. These results are averaged over 10 runs to avoid random peak performances, but in each individual run we observed the significant decrease in computation time. The strengthened primal/dual bound from Algorithm 1 is much tighter than the primal bound, more than halving the optimality gap for the larger instances.

5.3. Why is the dual formulation faster?

To understand the significant faster computation time of the dual formulation displayed in Table 3, we look at the dimensions (number of uncertain parameters, wait-and-see decisions, constraints on variables and constraints on uncertain parameters) for the case $N = 20$. We give the values of these dimensions in Table 3 using the same format as is in Table 1. We observe that the primal and dual formulation have the same characteristics,

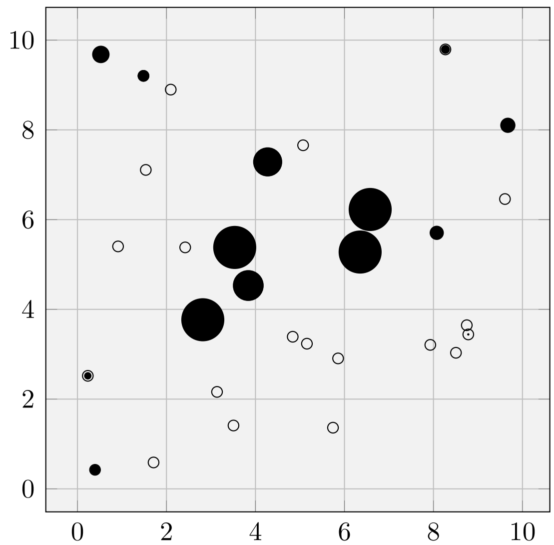

Figure 1 Stock allocation for an instance with 30 stores on the grid $[0,10]^2$. The filled dots have stock and the larger the dots are, the more stock is allocated. The open dots are stores that do not have any stock allocated.

Table 4 Comparing dimensions of variables, uncertainties and number of constraints in the primal and dual formulation for the lot-sizing instance with $N = 20$ stores.

	Primal formulation (2)	Dual formulation (3)
# uncertain parameters	20	21
# wait-and-see decisions	400	21
# constraints on variables	21	21
# constraints on uncertain parameter	21	401

except for the number of wait-and-see decisions and the number of constraints on the uncertain parameter in the uncertainty set. Given these values, we can explicitely calculate the number of affine constraints and the number of sign restrictions using the formulas from Table 2. The resulting number of constraints and sign restrictions are given in Table 5. We observe that the primal formulation (13) has about 50 times more affine constraints than the dual formulation (16). The dual formulation does have a lot more sign restrictions on its variables, but these are significantly simpler for solvers. To investigate the claim that the number of affine constraints are indeed the cause of the speedup we adapt the $N = 20$ instance from the network lot-sizing model (19). From Table 2 we see that increasing p, the number of affine constraints in the uncertainty set $\mathcal{U}$, leads to an increase of affine constraints in the dual formulation with affine policies. At the same time, the value of p

Table 5 **Comparing the number of affine constraints and sign restrictions in** (13) **and** (16) **for the lot-Sizing instance with** $N = 20$ **stores.**

	Primal formulation (13)	Dual formulation (16)
# affine constraints	8841	882
# sign restrictions	8841	16800

does not affect the number of affine constraints in the primal formulation. To increase p, we add nonreduntant constraints of the following type to the polyhedral description of $\mathcal{U}$:

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

where $S \subset \{1, \ldots, N\}$ is a random subset of size $\frac{1}{2}N$. The number of constraints p can be increased at will by adding more of these constraints. Note that increasing p also increases the total number of variables and the number of sign constraints, but these grow in more or less the same order of magnitude in both formulations. If we consider the case $N = 20$, then we find that the number of affine constraints in (13) and (16) is equal when the number of constraints in the uncertainty set $\mathcal{U}$ equals $p = 400$. Note that $p \geq 21$, since we need 21 constraints to describe the budget uncertainty set. The case with $p = 21$ is therefore just our original network lot-sizing problem (19). We measure the difference in computation time between the primal and the dual formulation by the quotient

$$\frac{\textit{Solver time for } (13)}{\textit{Solver time for } (16)}.$$

In Figure 2, we plotted this quotient for each random instance with $p \in \{21, 22, \ldots, 1000\}$ constraints in the uncertainty set. We already know from Table 3 that the dual formulation with affine policies solves the original instance three or more times faster than the primal formulation. If we start adding constraints, the computational advantage progressively decreases and after a point it dissapears.

6. Example 2: Facility Location Problem

The second example we consider is a facility location problem that has also been studied in Ardestani-Jaafari and Delage (2014) and Baron et al. (2011). Similar two-stage adaptive models can be found in Zeng and Zhao (2013). In our results we again observe a significant reduce in computational time required for solving the dualized formulation with affine

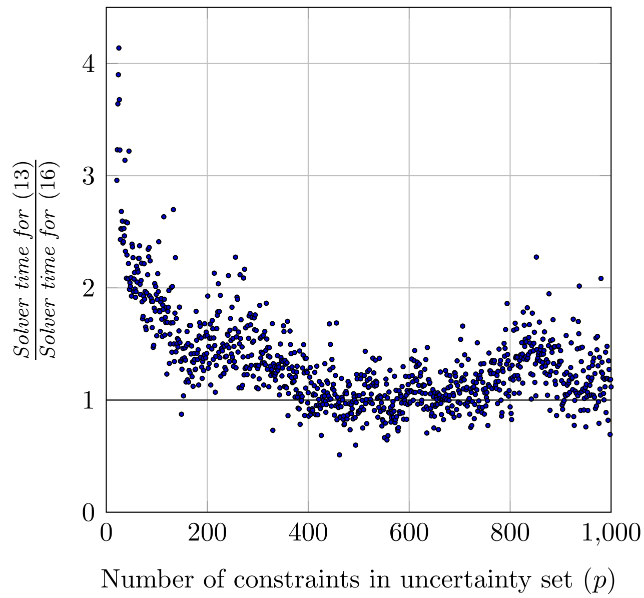

Figure 2 **The computation time of the primal formulation** (13) **divided by the time needed to solve the primal formulation** (13) **for** $N = 20$ **and various number of affine constraints in the uncertainty set** p**. For** $p = 400$ **both** (13) **and** (16) **have the same number of affine constraints. Values above the horizontal line at** 1 **indicate that the dual formulation is solved faster than the primal formulation and vice versa for values smaller than** 1**.**

policies over the primal formulation with affine policies. For this problem, however, the strenghtened bounds from Algorithm 1 only slightly improve the bounds obtained from the primal formulation.

6.1. Problem Setting

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

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

where Γ is our budget parameter[1]. There are two types of decisions in this model. First, strategic here-and-now decisions that have to be decided before the demand is known. We have a binary variable x_s to decide whether the facility at site s is opened and a

[1] In fact, Ardestani-Jaafari and Delage (2014) also consider negative values of the uncertainy parameter. It is not hard to see that these are nonbinding scenarios and we can therefore use this uncertainty set instead.

continuous variable p_s to set the capacity level at each opened facility site. Second, wait-and-see decisions y_{sc} on the production at facility s which is transported to customer c. Each unit of demand can generate a revenue of η. There are also several costs incurred for the various strategic and operations decisions. Opening a facility s has a fixed cost f_s and a cost of b_s per unit of capacity installed. The production of one unit at facility s has cost g_s and transporting the goods to customer c bears an additional cost h_{sc}. The goal is to maximize the total profit. This problem can be modelled as a two-stage adaptive optimization model, see Ardestani-Jaafari and Delage (2014):

$$
\begin{aligned}
\max_{t,\boldsymbol{x},\boldsymbol{p}} \quad & \alpha - \sum_{s\in\mathcal{S}} (b_s p_s + f_s x_s) \\
\text{s.t.} \quad & \forall \boldsymbol{\zeta} \in \mathcal{U}: \quad \exists \boldsymbol{y} \geq 0: \begin{cases} \sum_{s\in\mathcal{S}, c\in\mathcal{C}} (\eta - g_s - h_{sc})\, y_{sc} \geq \alpha \\ \sum_{c\in\mathcal{C}} y_{sc} \leq p_s \qquad \forall s \in \mathcal{S} \\ \sum_{s\in\mathcal{S}} y_{sc} \leq \bar{d}_c - \zeta_c \bar{d}_c \qquad \forall c \in \mathcal{C} \end{cases} \\
& \boldsymbol{p} \leq M\boldsymbol{x}, \ \boldsymbol{x} \in \{0,1\}^N.
\end{aligned}
\tag{21}
$$

Note that we have a maximization objective, but this can easily be turned into a minimization objective by the relation $\max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}) = -\min_{\boldsymbol{x}\in\mathcal{X}} (-f(\boldsymbol{x}))$ before applying Theorem 1.

6.2. Test Case and Numerical Results

We consider the same setting as in Ardestani-Jaafari and Delage (2014), which is based on the set-up of an earlier paper on robust facility location planning by Baron et al. (2011). We randomly generate C customers and S sites on a unit square. For the cost parameters we take $f_s = 50000$, $b_s = 0.1$, $g_s = 0.1$, $\eta = 1$. The nominal demand is drawn uniformly at random from $[17500, 22500]$ and $\hat{\zeta} = 0.15$. The transportation cost t_{ij} is just the Euclidean distance between two points i and j. We take $S = 10$ possible sites and $C \in \{10, 20, 30, 40, 50\}$. The cases with $C = 10$ and $C = 20$ are in Ardestani-Jaafari and Delage (2014) referred to as small and medium instances. For the larger instances the computational time vastly increased and they did not report results on the models with affine policies. We use the same computer and optimization software as mentioned in Section 5.

The results for various numbers of customers C and various percentage levels of uncertainty Γ are given in Table 6. We use the standard notion of budget uncertainty where a budget of 30% means that 30% of the uncertain prameters can be at their extreme value of $\hat{\zeta} = 0.15$. A graph indicating the location and the facilities that are opened for one case is given in Figure 3.

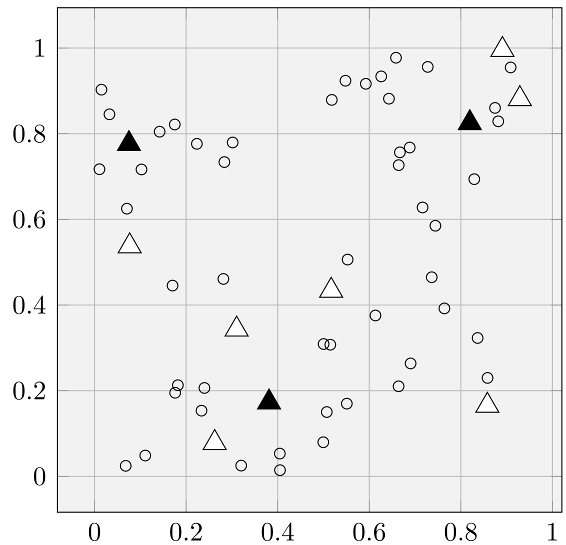

Figure 3 **Solution for one facility location instance with $S = 10$ possible sites and $C = 50$ customers on $[0.1]^2$. The uncertainty level is set at $\Gamma = 50\%$. Facility locations are indicated by triangles, customers by open circles. The filled triangles are the locations that are picked to be open.**

The most striking result is that the dual formulation with affine policies is again solved an order of magnitude faster than the primal formulation with affine policies. This holds especially true for the larger instances and larger values of Γ. We again look at the dimensions of the primal and the dual formulation using Table 1 for its dimensions and Table 2 for the different constraints. For the the case with $C = 50$ customers we present these results in Table 7 and Table 8.

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

If we take a look at the bounds we see they are very close to the objective value, which shows that the use of affine policies is nearly optimal. This observation was also made for the smaller instances in Ardestani-Jaafari and Delage (2014). For $\Gamma = 100\%$, the lower bound is the most far away from the objective value. This is surprising, as for this case (box uncertainty) we know that affine (in fact, static) policies are provably optimal (Ben-Tal et al. 2009, Theorem 14.2.4).

Table 6 Numerical results for facility location problem with affine policies. The percentages in the last columns depict the optimality gap derived from each upper bound compared to the objective value. All results are averaged over 5 runs.

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

Table 7 Comparing dimensions of variables, uncertainties and number of constraints in the primal and dual formulation for the facility location problem (21) with $C = 50$ customers.

	Primal formulation (2)	Dual formulation (3)
# uncertain parameters	50	61
# wait-and-see decisions	500	51
# constraints on variables	61	51
# constraints on uncertain parameter	51	501

Table 8 Comparing the number of affine constraints and sign restrictions in (13) and (16) for the facility location problem (21) with $C = 50$ customers.

	Primal formulation (13)	Dual formulation (16)
# affine constraints	28661	6222
# sign restrictions	28661	51000

7. Concluding Remarks

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

References

Ardestani-Jaafari, A., E. Delage. 2014. The value of flexibility in robust location-transportation problem *Les Cahiers du GERAD G–2014–83, GERAD, HEC Montréal.*

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

Baron, O., J. Milner, H. Naseraldin. 2011. Facility location: A robust optimization approach. *Production and Operations Management* **20** 772–785.

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

Ben-Tal, A., B. Do Chung, S. R. Mandala, T. Yao. 2011. Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation research part B: methodological* **45** 1177–1189.

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

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

Bertsimas, D., D. B. Brown, C. Caramanis. 2011. Theory and applications of robust optimization. *SIAM Review* **53** 464–501.

Bertsimas, D., C. Caramanis. 2010. Finite adaptability in multistage linear optimization. *Automatic Control, IEEE Transactions on* **55** 2751–2766.

Bertsimas, D., I. Dunning. 2014. Multistage robust mixed integer optimization with adaptive partitions. *Submitted for publication.*

Bertsimas, D., I. Dunning, M. Lubin. 2015. Reformulation versus cutting-planes for robust optimization. *Computational Management Science* 1–23.

Bertsimas, D., V. Goyal. 2010. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research* **35** 284–305.

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

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

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

Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52** 35–53.

Birge, J. R., F. Louveaux. 2011. *Introduction to stochastic programming*. Springer Science & Business Media.

Fischetti, M., M. Monaci. 2012. Cutting plane versus compact formulations for uncertain (integer) linear programs. *Mathematical Programming Computation* **4** 239–273.

Gabrel, V., M. Lacroix, C. Murat, N. Remli. 2014a. Robust location transportation problems under uncertain demands. *Discrete Applied Mathematics* **164** 100–111.

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

Gurobi Optimization, Inc. 2015. Gurobi optimizer reference manual. `http://www.gurobi.com`.

Hadjiyiannis, M., P. Goulart, D. Kuhn. 2011. A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*. 7386–7391.

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

Kali, P., S. W. Wallace. 1994. *Stochastic programming*. Springer.

Kuhn, D., W. Wiesemann, A. Georghiou. 2011. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* **130** 177–209.

Lubin, M., I. Dunning. 2015. Computing in operations research using julia. *INFORMS Journal on Computing* **27** 238–248.

Minoux, M. 2011. On 2-stage robust lp with rhs uncertainty: complexity results and applications. *Journal of Global Optimization* **49** 521–537.

Mutapcic, A., S. Boyd. 2009. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software* **24** 381–406.

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

Postek, K., D. Den Hertog. 2014. Multi-stage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *Submitted for publication.*

Shapiro, A., A. Nemirovski. 2005. On complexity of stochastic programming problems. *Continuous optimization*. Springer, 111–146.

Thiele, A., T. Terry, M. Epelman. 2009. Robust linear optimization with recourse. *Rapport technique* 4–37.

Wang, Q., J.-P. Watson, Y. Guan. 2013. Two-stage robust optimization for nk contingency-constrained unit commitment. *Power Systems, IEEE Transactions on* **28** 2366–2375.

Yin, Y., S. M. Madanat, X.-Y. Lu. 2009. Robust improvement schemes for road networks under demand uncertainty. *European Journal of Operational Research* **198** 470–479.

Zeng, B., L. Zhao. 2013. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters* **41** 457–461.

Zhao, L., B. Zeng. 2012. Robust unit commitment problem with demand response and wind energy. *Power and Energy Society General Meeting, 2012 IEEE*. 1–8.