Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets

Bram L. Gorissen[1,2], Aharon Ben-Tal[3], Hans Blanc[1], Dick den Hertog[1]

[1] *Department of Econometrics and Operations Research, Tilburg University.*
[2] *Corresponding author. E-mail:* `b.l.gorissen@tilburguniversity.edu.`
[3] *Department of Industrial Engineering and Management, Technion - Israel Institute of Technology.*
[3] *CentER Extramural Fellow, Tilburg University.*

Abstract

We propose a new way to derive tractable robust counterparts of a linear program by using the theory of Beck and Ben-Tal (2009) on the duality between the robust ("pessimistic") primal problem and its "optimistic" dual. First, we obtain a new *convex* reformulation of the dual problem of a robust linear program, and then show how to construct the primal robust solution from the dual optimal solution. Our result allows many new uncertainty regions to be considered. We give examples of tractable uncertainty regions that were previously intractable. The results are illustrated by solving a multi-item newsvendor problem. We also apply the new method to the globalized robust counterpart scheme and show its tractability.

1 Introduction

Robust Optimization (RO) is a paradigm for dealing with uncertain data in an optimization problem. Parts of RO originate from the seventies and eighties (Soyster 1974, Thuente 1980, Singh 1982, Kaul et al. 1986), but most of the existing theory and applications followed after new results in the late nineties (Ben-Tal and Nemirovski 1998, El Ghaoui and Lebret 1997). An extensive overview of RO is given in (Ben-Tal et al. 2009) and the survey (Bertsimas et al. 2011). The basic idea of RO is that constraints have to hold for all parameter realizations in some given uncertainty region.

Currently, two tractable methods to solve an RO problem can be distinguished. Both methods are applied constraint-wise, i.e. they reformulate individual constraints. The first method uses conic duality (e.g. used by Ben-Tal et al. (2009)), while the second method uses Fenchel duality (Ben-Tal et al. 2015). However, there are still some uncertainty sets for which both methods may not produce explicit tractable robust counterparts (RCs).

The purpose of this paper is to present a new method for robust linear programs, based on the result "primal worst equals dual best" (Ben-Tal et al. 2009), which gives tractable optimization problems for general convex uncertainty regions. In Section 3, we give examples of uncertainty sets where the new method generates explicit tractable robust counterparts, whereas the classical methods result in intractable RCs.

We also apply the new method to the *globalized robust counterpart* (GRC) model. In this model, there are two convex uncertainty regions, where the constraint must hold for uncertain events in the smaller uncertainty region, while the violation of the constraint for the events in the larger region is controlled via a convex distance function. We show that the GRC can be formulated as well as an ordinary robust linear program with a (different) convex uncertainty region, which implies that it can be solved efficiently with the method presented in this paper.

This is an author-created, un-copyedited version of a technical note published in Operations Research DOI:10.1287/opre.2014.1265.

2 A method for deriving a tractable dual of the Robust Counterpart

Consider the following Linear Conic Program (LCP):

$$\text{(LCP)} \quad \max_{\boldsymbol{x}\in\mathcal{K}} \{\boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i \quad \forall i \in I\},$$

where $\mathcal{K}$ is a closed convex cone and I is a finite index set. If $\mathcal{K}$ is the nonnegative orthant $\mathbb{R}^n_+$, (LCP) is an LP in canonical form. Other common choices for $\mathcal{K}$ are the second-order cone and the semidefinite cone. This setting is more general than the title of this paper indicates, but we will only allow uncertainty in the linear constraints and want to avoid confusion with robust conic optimization. We will use the prefix D for the dual, O for the optimistic counterpart, and R for the robust counterpart. The dual of (LCP) is given by:

$$\text{(D-LCP)} \quad \min_{\boldsymbol{y}\ge\boldsymbol{0}} \{\boldsymbol{b}^\top \boldsymbol{y} : \sum_{i\in I} y_i \boldsymbol{a}_i - \boldsymbol{c} \in \mathcal{K}^*\},$$

where $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$ and I is a finite index set. Assume that $\boldsymbol{a}_i$ are uncertain, but known to reside in some convex compact uncertainty region $\mathcal{U}_i = \{\boldsymbol{a}_i : f_{ik}(\boldsymbol{a}_i) \le 0 \quad \forall k \in K\}$, where f_{ik} are given closed proper convex functions and K is a finite index set. The robust counterpart of (LCP) is given by:

$$\begin{aligned} \text{(R-LCP)} \quad \max_{\boldsymbol{x}\in\mathcal{K}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i \qquad \forall \boldsymbol{a}_i : f_{ik}(\boldsymbol{a}_i) \le 0 \quad \forall i \in I \quad \forall k \in K, \end{aligned} \tag{1}$$

(R-LCP) is a semi-infinite convex optimization problem since it has a linear objective, $\boldsymbol{x}$ is finite dimensional, and the feasible region is the intersection of infinitely many half spaces. Since (R-LCP) is convex, a local optimum is also a global optimum. Numerical methods that find this optimum cannot be applied because of the semi-infinite representation. Current RO rewrites constraint (1) to a finite set of constraints, which works for a limited set of functions f_{ik}. The optimistic counterpart of (D-LCP) is given by:

$$\text{(OD-LCP)} \quad \min_{\boldsymbol{y}\ge\boldsymbol{0}} \{\boldsymbol{b}^\top \boldsymbol{y} : \exists \boldsymbol{a}_i,\ f_{ik}(\boldsymbol{a}_i) \le 0\,,\ \forall i \in I\ \forall k \in K,\ \sum_{i\in I} y_i \boldsymbol{a}_i - \boldsymbol{c} \in \mathcal{K}^*\}.$$

A result by Beck and Ben-Tal (2009) is that (OD-LCP), which is optimistic since it has to hold for a single $\boldsymbol{a}_i$, is a dual problem of (R-LCP), which has to hold for all $\boldsymbol{a}_i$. Moreover, the values of (R-LCP) and (OD-LCP) are equal if (OD-LCP) is bounded and satisfies the Slater condition. Less general but similar results can be found in (Falk 1976, Römer 2010, Soyster 1974, Thuente 1980). For $\mathcal{K} = \mathbb{R}^n_+$, (OD-LCP) is called a Generalized LP (GLP) (Dantzig 1963, p. 434). It contains the product of variables $y_i\boldsymbol{a}_i$ and is in general nonconvex. However, it will be shown that (OD-LCP) can be converted to a convex program.

Dantzig mentions substituting $\boldsymbol{v}_i = y_i\boldsymbol{a}_i$ and multiplying the constraints containing f_{ik} with y_i as a solution approach to GLPs (Dantzig 1963, p. 434), which has already been applied to the dual of LPs with polyhedral uncertainty by Römer (2010). When we apply this to (OD-LCP), we get the following convex optimization problem:

$$\begin{aligned} \text{(COD-LCP)} \quad \min_{\boldsymbol{y}\ge\boldsymbol{0},\boldsymbol{v}_i} \quad & \boldsymbol{b}^\top \boldsymbol{y} \\ \text{s.t.} \quad & \sum_{i\in I} \boldsymbol{v}_i - \boldsymbol{c} \in \mathcal{K}^* && (2) \\ & y_i f_{ik}\left(\frac{\boldsymbol{v}_i}{y_i}\right) \le 0\,, \qquad \forall i \in I \quad \forall k \in K, && (3) \end{aligned}$$

where $0f_{ik}\left(\boldsymbol{v}_i/0\right) = \lim_{y_i \downarrow 0} y_i f_{ik}\left(\boldsymbol{v}_i/y_i\right)$ is the recession function of f. (COD-LCP) is indeed a convex problem, since the perspective function $g_{ik}(\boldsymbol{v}_i, y_i) := y_i f_{ik}\left(\boldsymbol{v}_i/y_i\right)$ is convex on $\mathbb{R}^n \times \mathbb{R}_+$. Here is a short proof of the convexity of $y_i f_{ik}(\boldsymbol{v}_i/y_i)$ on $\mathbb{R}^n \times \mathbb{R}_+ \backslash \{0\}$ that uses convex analysis:

$$g_{ik}(\boldsymbol{v}_i, y_i) = y_i f_{ik}\left(\frac{\boldsymbol{v}_i}{y_i}\right) = y_i f_{ik}^{**}\left(\frac{\boldsymbol{v}_i}{y_i}\right) = y_i \sup_{\boldsymbol{x}} \left\{\frac{\boldsymbol{v}_i^\top}{y_i}\ \boldsymbol{x} - f_{ik}^*(\boldsymbol{x})\right\} = \sup_{\boldsymbol{x}} \left\{\boldsymbol{v}_i^\top\ \boldsymbol{x} - y_i f_{ik}^*(\boldsymbol{x})\right\},$$

from which it follows that g_{ik} is jointly convex since it is the pointwise supremum of functions which are linear in $\boldsymbol{v}_i$ and y_i.

While (R-LCP) is difficult to solve because it has an infinite number of constraints, (COD-LCP) does not have "for all" constraints. For some popular choices of f_{ik} for which an exact reformulation of (R-LCP) is known, (COD-LCP) is at most as difficult to solve as (R-LCP). For instance, when the uncertainty region is polyhedral, (R-LCP) can be reformulated as an LP, and (COD-LCP) is also an LP. When the uncertainty region is an ellipsoid, (R-LCP) can be reformulated as a conic quadratic program, and (COD-LCP) is also a conic quadratic program.

Dantzig notes that (OD-LCP) and (COD-LCP) are equivalent only when $\boldsymbol{v}_i \neq \boldsymbol{0}$ is not possible if $y_i = 0$, since otherwise $\boldsymbol{v}_i = y_i \boldsymbol{a}_i$ does not hold. We call this the *substitution equivalence condition.* The following lemma shows that this condition is automatically satisfied:

Lemma 1 *Assume that the uncertainty region is bounded. Then* (3) *enforces the substitution equivalence condition.*

Proof. Let $i \in I$ and let $y_i = 0$. From the definition of $0f_{ik}\left(\boldsymbol{v}_i/0\right)$, it is clear that $\boldsymbol{v}_i = \boldsymbol{0}$ is feasible for (3). It remains to show that a nonzero $\boldsymbol{v}_i^*$ is infeasible. Assume to the contrary that $\boldsymbol{v}_i^* \neq \boldsymbol{0}$ is feasible. Let us first construct two points where $g_{ik}(\boldsymbol{v}_i, y_i) \leq 0$.

The first point is $(c\boldsymbol{v}_i^*, 0)$, $c > 0$, and the second is $(2a_i, 2)$ where $a_i \in \mathcal{U}_i$. Indeed, for $k \in K$, $g_{ik}(c\boldsymbol{v}_i^*, 0) = \lim_{y_i \downarrow 0} y_i f_{ik}\left(c\boldsymbol{v}_i^*/y_i\right) = c \lim_{y_i \downarrow 0}(y_i/c) f_{ik}\left(\boldsymbol{v}_i^*/(y_i/c)\right) = c g_{ik}(\boldsymbol{v}_i^*, 0) \leq 0$ and clearly also $g_{ik}\left(2\boldsymbol{a}_i, 2\right) \leq 0$ for an arbitrary $\boldsymbol{a}_i \in \mathcal{U}_i$.

By convexity, we have $g_{ik}\left(\lambda \boldsymbol{v}_i^1 + (1-\lambda)\boldsymbol{v}_i^2, \lambda y_i^1 + (1-\lambda) y_i^2\right) \leq \lambda g_{ik}\left(\boldsymbol{v}_i^1, y_i^1\right) + (1-\lambda) g_{ik}\left(\boldsymbol{v}_i^2, y_i^2\right)$. In particular, for $\lambda = 0.5$ and the above two points, we get $g_{ik}((c\boldsymbol{v}_i^* + 2\boldsymbol{a}_i)/2, 1) \leq 0$ for all $k \in K$. This implies that $(c/2)\boldsymbol{v}_i^* + \boldsymbol{a}_i$ is in $\mathcal{U}_i$ for all $c > 0$. So, the uncertainty region recedes in the direction of $\boldsymbol{v}_i^*$, contradicting boundedness. ■

In practice it is often necessary to have the primal robust solution $\boldsymbol{x}$ of (R-LCP), instead of the solution of (COD-LCP). The following theorem shows how $\boldsymbol{x}$ can be recovered from an optimal solution of (COD-LCP).

Theorem 1 *Assume that (COD-LCP) is bounded and satisfies the Slater condition. A KKT vector of constraint* (2) *corresponds to an optimal solution $\boldsymbol{x}$ of (R-LCP).*

Proof. First, we show that the dual variables associated with constraint (2) are the optimization variables of (R-LCP). The Lagrangian of (COD-LCP) is given by:

$$L(\boldsymbol{y}, v, \boldsymbol{x}) = \begin{cases} \boldsymbol{b}^\top\ \boldsymbol{y} + \boldsymbol{x}^\top\ \left(\boldsymbol{c} - \sum_{i \in I} \boldsymbol{v}_i\right) & \text{if } y_i f_{ik}\left(\frac{\boldsymbol{v}_i}{y_i}\right) \leq 0, \qquad \forall i \in I \quad \forall k \in K \\ \infty & \text{otherwise,} \end{cases}$$

and hence, (R-LCP) is given by:

$$\begin{aligned}
&\max_{\boldsymbol{x}\in\mathcal{K}} \min_{\boldsymbol{y}\geq\boldsymbol{0},v} L(\boldsymbol{y},v,\boldsymbol{x}) \\
= \;& \max_{\boldsymbol{x}\in\mathcal{K}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \min_{\boldsymbol{y}\geq\boldsymbol{0},v} \left\{ \boldsymbol{b}^\top \boldsymbol{y} - \sum_{i\in I} \boldsymbol{v}_i^\top \boldsymbol{x} : y_i f_{ik}\left(\frac{\boldsymbol{v}_i}{y_i}\right) \leq 0\,, \quad \forall i\in I\ \forall k\in K \right\}\right\} \\
= \;& \max_{\boldsymbol{x}\in\mathcal{K}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} + \min_{\boldsymbol{y}\geq\boldsymbol{0},a} \left\{ \sum_{i\in I} y_i \left(b_i - \boldsymbol{a}_i^\top \boldsymbol{x}\right) : f_{ik}\left(\boldsymbol{a}_i\right) \leq 0 \quad \forall i\in I\ \forall k\in K \right\}\right\} \\
= \;& \max_{\boldsymbol{x}\in\mathcal{K}} \left\{ \boldsymbol{c}^\top \boldsymbol{x} : \boldsymbol{a}_i^\top \boldsymbol{x} \leq b_i \quad \forall \boldsymbol{a}_i : f_{ik}(\boldsymbol{a}_i) \leq 0\,, \quad \forall i\in I\ \forall k\in K \right\},
\end{aligned}$$

where in the second equality the substitution $\boldsymbol{a}_i = \boldsymbol{v}_i/y_i$ is made. If an optimal value is attained at $y_i = 0$ (and consequently at $\boldsymbol{v}_i = \boldsymbol{0}$) before the substitution, then any feasible $\boldsymbol{a}_i$ is optimal after the substitution. The problem in the last equality is indeed (R-LCP).

Since (COD-LCP) is bounded and satisfies the Slater condition, a KKT vector exists (Rockafellar 1970, Theorem 28.2). An optimal $\boldsymbol{x}$ of (R-LCP) is equal to a KKT vector (Rockafellar 1970, Corollary 28.4.1). ■

This theorem is useful in practice because many solvers can output a KKT vector. There is also another way to obtain a solution of (R-LCP), similar to the method mentioned used by Soyster (1974). The idea is to use the "dual best $\boldsymbol{a}_i$" as the "primal worst $\boldsymbol{a}_i$": translate a solution of (COD-LCP) to a solution of (OD-LCP), then fix the variables $\boldsymbol{a}_i$, remove the constraints on $\boldsymbol{a}_i$, and dualize that problem with respect to y_i. The result is a problem similar to (LCP), where the vectors $\boldsymbol{a}_i$ have been replaced with "worst case" $\boldsymbol{a}_i$. We call this problem (M-LCP). This method only works if (COD-LCP) has a unique optimal $\boldsymbol{a}_i$, and if (M-LCP) has a unique optimal $\boldsymbol{x}$ (Soyster 1974). Then, the value of (M-LCP) equals the value of (OD-LCP), and $\boldsymbol{x}$ is both feasible and optimal for (R-LCP).

The main advantage of our method is the fact that we use the perspectives of the original functions that define the uncertainty region, while existing methods in RO use the perspective of the conjugate functions to reformulate the "for all" constraints in (R-LCP) (Ben-Tal et al. 2015); the latter may not result in closed-form formulations for many uncertainty regions. We give examples of the above in Section 3.

There may be an additional computational advantage as the number of variables and constraints in (COD-LCP) can be smaller, compared to results obtained with existing RO techniques. For example, the latter may require an explicit conic representation (e.g. see (Ben-Tal et al. 2009, Theorem 1.3.4)) which can significantly increase the number of variables and constraints.

Our method also has two disadvantages that is inherent to solving the *dual* problem. First, it cannot directly be applied to problems with integer variables (however, it can be used to solve LP relaxations such as those needed in cutting plane and branch & bound methods). Second, the primal solution has to be recovered from the KKT vector. This means that the dual problem has to be solved to high accuracy for otherwise the accuracy of the *primal* solution may suffer. However, the actual effect on the optimal primal objective function can be assessed by the duality gap. We show this in our numerical example in Section 5.

3 New tractable uncertainty regions

In this section we present three examples of uncertainty regions for which the robust counterpart cannot be obtained using the traditional approach:

1. The first example is given by problems in which several scenarios for the parameters can be distinguished, but the probabilities on these scenarios are not known. Suppose these unknown

probabilities can be estimated based on historical data, and an optimization problem has a constraint involving these probabilities. An example of this is a constraint on expected value. For such problems, a wide class of uncertainty regions is given in terms of the distance between the real probability vector $\boldsymbol{p}$ and a historical estimate $\hat{\boldsymbol{p}}$, both indexed by the scenario s from a finite scenario set S:

$$\mathcal{U}_i = \left\{ \boldsymbol{p} : \boldsymbol{p} \geq \boldsymbol{0}, \sum_{s \in S} p^{(s)} = 1, d(\boldsymbol{p}, \hat{\boldsymbol{p}}) \leq \rho \right\}, \tag{4}$$

where d is the distance measure and ρ is the level of uncertainty. Note that the constraint $\sum_{s \in S} p^{(s)} = 1$ is not necessary for the following results to hold, so $\boldsymbol{p}$ does not need to be a probability vector. We consider several classes of distance measures:

(a) The first class of distance measures that contains previously intractable cases is ϕ–divergence, which for a convex function ϕ that satisfies $\phi(1) = 0$ is defined by $d(\boldsymbol{p}, \hat{\boldsymbol{p}}) = \sum_{s \in S} \hat{p}^{(s)} \phi \left(p^{(s)} / \hat{p}^{(s)} \right)$. Ben-Tal et al. (2013) show how to choose ρ in (4) based on historical observations, and give tractable robust counterparts for several choices of ϕ. One example for which their method does not give a tractable reformulation is the Matusita distance (Matusita 1967), where $\phi(t) = |1 - t^{\alpha}|^{1/\alpha}$ for given $\alpha \in (0, 1)$.

(b) The second class of distance measures is based on the Bregman distance, which is given by (Bregman 1967):

$$d(\boldsymbol{p}, \hat{\boldsymbol{p}}) = g(\boldsymbol{p}) - g(\hat{\boldsymbol{p}}) - (\nabla g(\hat{\boldsymbol{p}}))^{\top} (\boldsymbol{p} - \hat{\boldsymbol{p}}),$$

where g is real-valued, continuously-differentiable and strictly convex on the set of probability vectors. The Bregman distance is convex in its first argument. Previously, uncertainty regions were intractable for many choices of g, while with our results any g gives a tractable optimistic counterpart.

(c) The third class of distance measures is the Rényi divergence (Rényi 1961):

$$d(\boldsymbol{p}, \hat{\boldsymbol{p}}) = \frac{1}{\alpha - 1} \log \sum_{i \in I} \left(\hat{p}_i^{(s)} \right)^{\alpha} \left(p_i^{(s)} \right)^{1-\alpha},$$

where $\alpha > 0$ and $\alpha \neq 1$. After some rewriting, an uncertainty region based on this distance measure can also be reformulated using Fenchel duality (Ben-Tal et al. 2015). However, the rewriting is not always possible, e.g. when this divergence measure is clustered with other distance measures (Banerjee et al. 2005), while our result can then still be applied.

2. The second example of new tractable uncertainty regions is when the uncertainty region contains an uncertain parameter $\boldsymbol{B}_{ij}$:

$$\mathcal{U}_i = \left\{ (\boldsymbol{a}_i, \boldsymbol{\zeta}_i) : \exists \boldsymbol{B}_{ij},\ \boldsymbol{\zeta}_i \geq \boldsymbol{0},\ \boldsymbol{a}_i = \boldsymbol{a}_i^0 + \sum_j \zeta_{ij} \boldsymbol{B}_{ij},\ g_{ijk}(\boldsymbol{B}_{ij}) \leq 0,\ h_{ik}(\boldsymbol{\zeta}_i) \leq 0, \quad \forall j \in J, k \in K \right\},$$

where g_{ijk} and h_{ik} are convex functions and $\boldsymbol{B}_{ij}, \boldsymbol{a}_i^0$ are vectors. The same substitution we applied to (OD-CLP), can also be applied to this uncertainty region. Let $\boldsymbol{v}_{ij} = \zeta_{ij} \boldsymbol{B}_{ij}$. The uncertainty region $\mathcal{U}_i$ can be rewritten as:

$$\mathcal{U}_i = \left\{ (\boldsymbol{a}_i, \boldsymbol{\zeta}_i) : \exists \boldsymbol{v}_{ij},\ \boldsymbol{\zeta}_i > \boldsymbol{0},\ \boldsymbol{a}_i = \boldsymbol{a}_i^0 + \sum_j \boldsymbol{v}_{ij},\ \zeta_{ij} g_{ijk}(\boldsymbol{v}_{ij} / \zeta_{ij}) \leq 0,\ h_{ik}(\boldsymbol{\zeta}_i) \leq 0, \quad \forall j \in J, k \in K \right\},$$

which is convex, and hence, leads to a tractable optimistic counterpart. We mention three cases where this uncertainty region appears:

(a) First, it appears in factor models where the parameter $\boldsymbol{a}_i$ is estimated as $\boldsymbol{a}_i^0 + \sum_j \zeta_{ij} \boldsymbol{B}_{ij}$ with uncertainty in both the factors $\boldsymbol{\zeta}_i$ and the coefficients $\boldsymbol{B}_{ij}$. For example in finance, the return of an asset can be approximated by $\mu + V^\top f$, where f are the factors that drive the market (Goldfarb and Iyengar 2003).

(b) Second, it appears in a constraint containing the steady-state distributions of a Markov chain, where the transition probabilities are uncertain. The uncertainty region then looks as follows:

$$\mathcal{U}_i = \left\{ \boldsymbol{\pi} \in \mathbb{R}^n_+ : \boldsymbol{e}^\top \boldsymbol{\pi} = 1, \ \sum_{j \in J} \pi_j \boldsymbol{B}_j = \boldsymbol{\pi}, \ g_{jk}(\boldsymbol{B}_j) \leq 0 \,, \quad \forall j \in J, k \in K \right\},$$

where $\boldsymbol{B}_j$ are the columns of the matrix with transition probabilities. Markov chains with column-wise uncertainty in the transition matrix were also considered by Blanc and den Hertog (2008).

(c) Third, it appears in a constraint on the next time period probability vector $\boldsymbol{p}_i$ of a Markov chain when there is uncertainty both in the transition matrix and in the current state:

$$\mathcal{U}_i = \left\{ \boldsymbol{p}_i \in \mathbb{R}^n_+ : \boldsymbol{p}_i = \sum_{j \in J} (\boldsymbol{p}_i^0)_j \boldsymbol{B}_j, \ g_{jk}(\boldsymbol{B}_j) \leq 0, \ h_k(\boldsymbol{p}_i^0) \leq 0 \,, \quad \forall j \in J, k \in K \right\},$$

where $\boldsymbol{B}_j$ are the columns of the transition matrix and $\boldsymbol{p}_i^0$ is the current probability vector.

3. The third example of new tractable uncertainty regions is illustrated by the following robust constraint:

$$\boldsymbol{a}_i^\top \boldsymbol{x} + \sum_{j \in J} h_{ij}(a_{ij}) x_j \leq b_i \qquad \forall \boldsymbol{a}_i : ||\boldsymbol{a}_i||_\infty \leq 1,$$

where the functions h_{ij} are convex. For many choices of h_{ij} this constraint is not tractable. To show that it can be solved with our method, we first move the nonlinearity to the uncertainty region:

$$\boldsymbol{a}_i^\top \boldsymbol{x} + \sum_{j \in J} d_{ij} x_j \leq b_i \qquad \forall (\boldsymbol{a}_i, \boldsymbol{d}_i) : ||\boldsymbol{a}_i||_\infty \leq 1, \quad h_{ij}(a_{ij}) = d_{ij} \quad \forall j \in J,$$

and then obtain an equivalent constraint by taking the convex hull of the uncertainty region:

$$\begin{aligned} \boldsymbol{a}_i^\top \boldsymbol{x} + \sum_{j \in J} d_{ij} x_j \leq b_i \qquad \forall (\boldsymbol{a}_i, \boldsymbol{d}_i) : ||\boldsymbol{a}_i||_\infty \leq 1, \quad h_{ij}(a_{ij}) \leq d_{ij}, \\ 2 d_{ij} \leq h_{ij}(1)(a_{ij} + 1) - h_{ij}(-1)(a_{ij} - 1) \quad \forall j \in J. \end{aligned}$$

This transformation has also been applied by Ben-Tal et al. (2015, p. 20), but they require a closed form for the convex conjugate of h_{ij} to reformulate this constraint. With our method, this linear constraint with a convex uncertainty region is tractable for any convex h_{ij}.

4 Globalized Robust Counterpart

A robust constraint holds for all realizations of the uncertain parameters in the uncertainty region. If this region contains all "physically possible" events, the RC may result in a very pessimistic solution. Ben-Tal et al. (2006) proposes the *globalized robust counterpart* (GRC) model, which may be used to reduce the conservatism of the RC (also see (Ben-Tal et al. 2009, Chapter 3)). Here we use a slightly modified version. Let $\mathcal{U}_i = \{\boldsymbol{a}_i : f_{ik}(\boldsymbol{a}_i) \leq 0 \,, \quad \forall k \in K\}$ be the set of "physically possible" realizations, and let a smaller set $\mathcal{U}_i' = \{\boldsymbol{a}_i : g_{ik}(\boldsymbol{a}_i) \le 0\,, \quad \forall k \in K\} \subset \mathcal{U}_i$ contain the "normal range" of realizations. We define the GRC as:

$$\boldsymbol{a}_i^\top \boldsymbol{x} \le b_i + \min_{\boldsymbol{a}_i' \in \mathcal{U}_i'} \{h_i(\boldsymbol{a}_i, \boldsymbol{a}_i')\} \qquad \forall \boldsymbol{a}_i \in \mathcal{U}_i, \tag{5}$$

where h_i is a nonnegative jointly convex distance-like function for which $h_i(\boldsymbol{a}_i', \boldsymbol{a}_i') = 0$ for all $\boldsymbol{a}_i'$ in $\mathcal{U}_i'$. Examples are norms and ϕ–divergence measures. The second term on the right-hand side of (5) expresses the allowable violation of the constraint and is equal to 0 if $\boldsymbol{a}_i$ is in the smaller set $\mathcal{U}_i'$.

We now show that (5) can be reformulated as a linear constraint with a convex uncertainty region. Constraint (5) is equivalent to:

$$\boldsymbol{a}_i^\top \boldsymbol{x} \le b_i + d_i \qquad \forall (\boldsymbol{a}_i, d_i) : f_{ik}(\boldsymbol{a}_i) \le 0 \quad \forall k \in K, \quad d_i = \min_{\boldsymbol{a}_i' \in \mathcal{U}_i'} \{h_i(\boldsymbol{a}_i, \boldsymbol{a}_i')\},$$

which in turn is equivalent to:

$$\boldsymbol{a}_i^\top \boldsymbol{x} \le b_i + d_i \qquad \forall (\boldsymbol{a}_i, \boldsymbol{a}_i', d_i) : f_{ik}(\boldsymbol{a}_i) \le 0 \quad \forall k \in K, \quad g_{ik}(\boldsymbol{a}_i') \le 0\,, \quad \forall k \in K, \quad d_i \ge h_i(\boldsymbol{a}_i, \boldsymbol{a}_i').$$

This is indeed a linear constraint with a convex uncertainty region. We will now show how the GRC can be formulated. Consider the following Globalized Robust program:

$$\begin{aligned} \text{(R-GRC)} \quad & \max_{\boldsymbol{x} \in \mathcal{K}} \quad \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{s.t.} \quad \boldsymbol{a}_i^\top \boldsymbol{x} \le b_i + d_i \quad \forall (\boldsymbol{a}_i, \boldsymbol{a}_i', d_i) : f_{ik}(\boldsymbol{a}_i) \le 0\,, \quad \forall k \in K, \\ & \qquad\qquad g_{ik}(\boldsymbol{a}_i') \le 0\,, \quad \forall k \in K \quad d_i \ge h_i(\boldsymbol{a}_i, \boldsymbol{a}_i') \quad \forall i \in I, \end{aligned}$$

whose optimistic dual is:

$$\begin{aligned} \text{(OD-GRC)} \quad & \min_{\boldsymbol{y} \ge \boldsymbol{0}, \boldsymbol{a}_i, \boldsymbol{a}_i', \boldsymbol{d}} \quad \boldsymbol{b}^\top \boldsymbol{y} + \boldsymbol{d}^\top \boldsymbol{y} \\ & \text{s.t.} \quad \sum_{i \in I} y_i \boldsymbol{a}_i - \boldsymbol{c} \in \mathcal{K}^* \\ & \qquad f_{ik}(\boldsymbol{a}_i) \le 0 \qquad \forall i \in I \quad \forall k \in K && (6) \\ & \qquad g_{ik}(\boldsymbol{a}_i') \le 0 \qquad \forall i \in I \quad \forall k \in K && (7) \\ & \qquad h_i(\boldsymbol{a}_i, \boldsymbol{a}_i') \le d_i \qquad \forall i \in I. && (8) \end{aligned}$$

We substitute $\boldsymbol{v}_i = y_i \boldsymbol{a}_i$, $\boldsymbol{v}_i' = y_i \boldsymbol{a}_i'$ and $w_i = y_i d_i$ and multiply constraints (6)–(8) with y_i:

$$\begin{aligned} \text{(COD-GRC)} \quad & \min_{\boldsymbol{y} \ge \boldsymbol{0}, \boldsymbol{v}_i, \boldsymbol{v}_i', \boldsymbol{w}} \quad \boldsymbol{b}^\top \boldsymbol{y} + \sum_{i \in I} w_i \\ & \text{s.t.} \quad \sum_{i \in I} \boldsymbol{v}_i - \boldsymbol{c} \in \mathcal{K}^* \\ & \qquad y_i f_{ik}\left(\frac{\boldsymbol{v}_i}{y_i}\right) \le 0 \qquad \forall i \in I \quad \forall k \in K \\ & \qquad y_i g_{ik}\left(\frac{\boldsymbol{v}_i'}{y_i}\right) \le 0 \qquad \forall i \in I \quad \forall k \in K \\ & \qquad y_i h_i\left(\frac{\boldsymbol{v}_i}{y_i}, \frac{\boldsymbol{v}_i{}'}{y_i}\right) \le w_i \qquad \forall i \in I. && (9) \end{aligned}$$

Note that the product $y_i \boldsymbol{a}_i'$ does not appear in (OD-GRC), but that the substitution $\boldsymbol{v}_i' = y_i \boldsymbol{a}_i'$ is still necessary to make (COD-GRC) convex. From the objective and (9) it is clear that the substitution equivalence condition holds for w_i even though the uncertainty region is not bounded in $\boldsymbol{d}$. For the tractability of (COD-GRC), all results regarding the functions that define the uncertainty region $\mathcal{U}_i$ also apply to the functions that define $\mathcal{U}'_i$. For example, when h_i is a ϕ–divergence measure, constraint (8) is given by:

$$\sum_{j\in J} a_{ij}\phi\left(a'_{ij}/a_{ij}\right) \leq d_i.$$

Constraint (9) then contains the perspective of ϕ, which is just:

$$\sum_{j\in J} v_{ij}\phi\left(v'_{ij}/v_{ij}\right) \leq w_i.$$

When in (8), h_i is a norm, i.e., $||\boldsymbol{a}_i - \boldsymbol{a}'_i||$, constraint (9) contains the same norm: $||\boldsymbol{v}_i - \boldsymbol{v}'_i|| \leq w_i$.

5 Multi-item newsvendor example

We demonstrate our new method on a robust LP with a convex uncertainty region that currently cannot be solved with other methods. To obtain this problem, we take a slightly modified version of the multi-item newsvendor problem described by Ben-Tal et al. (2013). There are 12 items indexed by i, and each has its own ordering cost c_i, selling price v_i, salvage price r_i, and unsatisfied demand loss l_i. So, when the order quantity Q_i is less than the demand d_i, the profit equals $v_iQ_i + l_i(Q_i - d_i) - c_iQ_i$, and $v_id_i + r_i(Q_i - d_i) - c_iQ_i$ otherwise. If $r_i \leq v_i + l_i$, this profit is concave piecewise linear in the decision variable Q_i. In practice the demand is not known, but for every item i we can define scenarios s in a scenario set S which occur with probability $p_i^{(s)}$, independently of other items, resulting in a demand of $d_i^{(s)}$. The goal is to determine Q_i such that the total ordering cost is minimized under the constraint that the expected profit is at least γ. This can be formulated as a robust LP as follows:

$$\begin{aligned}
\text{(R-NV)} \quad \min_{\boldsymbol{Q}\geq\boldsymbol{0},\boldsymbol{u}} \quad & \sum_{i\in\mathcal{I}} c_iQ_i \\
\text{s.t.} \quad & \sum_{i\in\mathcal{I}}\sum_{s\in S} p_i^{(s)}u_i^{(s)} \geq \gamma \\
& \forall \boldsymbol{p}_i \geq \boldsymbol{0} : \sum_{s\in S} p_i^{(s)} = 1, \sum_{s\in S}\left|\left(\hat{p}_i^{(s)}\right)^\alpha - \left(p_i^{(s)}\right)^\alpha\right|^{1/\alpha} \leq \rho, \quad \forall i\in\mathcal{I} \qquad (10) \\
& u_i^{(s)} + (c_i - r_i)\,Q_i \leq d_i^{(s)}\,(v_i - r_i) \qquad \forall i\in\mathcal{I} \quad \forall s\in S \\
& u_i^{(s)} + (c_i - v_i - l_i)\,Q_i \leq -d_i^{(s)}l_i \qquad \forall i\in\mathcal{I} \quad \forall s\in S,
\end{aligned}$$

where $u_i^{(s)}$ denotes the profit for item i in scenario s, and the uncertainty region is based on the Matusita distance with α in $(0,1)$. (R-NV) cannot be solved with current methods, because the uncertainty region in (10) does not have a conic quadratic representation and (10) cannot be reformulated to finitely many constraints with closed form functions using Fenchel duality to the best of our knowledge (Ben-Tal et al. 2015, Table 1). The optimistic dual of (R-NV) is given by:

$$\begin{aligned}
\text{(OD-NV)} \quad \max_{x\geq 0,\boldsymbol{y}\leq\boldsymbol{0},\boldsymbol{z}\leq\boldsymbol{0},p\geq 0} \quad & \gamma x + \sum_{i\in\mathcal{I},s\in S} d_i^{(s)}\left((v_i - r_i)\,y_{is} - l_iz_{is}\right) \\
\text{s.t.} \quad & p_i^{(s)}x + y_{is} + z_{is} = 0 \qquad \forall i\in\mathcal{I} \quad \forall s\in S \\
& \sum_{s\in S}\left\{(c_i - r_i)\,y_{is} + (c_i - v_i - l_i)\,z_{is}\right\} \leq c_i \qquad \forall i\in\mathcal{I} \\
& \sum_{s\in S} p_i^{(s)} = 1 \qquad \forall i\in\mathcal{I} \qquad (11) \\
& \sum_{s\in S}\left|\left(\hat{p}_i^{(s)}\right)^\alpha - \left(p_i^{(s)}\right)^\alpha\right|^{1/\alpha} \leq \rho \qquad \forall i\in\mathcal{I}. \qquad (12)
\end{aligned}$$

After substituting $w_{is} = p_i^{(s)} x$ and multiplying (11) and (12) with x, the convex reformulation becomes:

$$\text{(COD-NV)} \quad \max_{x \geq 0, \boldsymbol{y} \leq \boldsymbol{0}, \boldsymbol{z} \leq \boldsymbol{0}, w \geq 0} \quad \gamma x + \sum_{i \in \mathcal{I}, s \in S} d_i^{(s)} \left((v_i - r_i) y_{is} - l_i z_{is} \right)$$

$$\text{s.t.} \quad w_{is} + y_{is} + z_{is} = 0 \qquad \forall i \in \mathcal{I} \quad \forall s \in S \tag{13}$$

$$\sum_{s \in S} \left\{ (c_i - r_i) y_{is} + (c_i - v_i - l_i) z_{is} \right\} \leq c_i \qquad \forall i \in \mathcal{I} \tag{14}$$

$$\sum_{s \in S} w_{is} = x \qquad \forall i \in \mathcal{I}$$

$$\sum_{s \in S} \left| \left(\hat{p}_i^{(s)} x \right)^{\alpha} - w_{is}^{\alpha} \right|^{1/\alpha} \leq \rho x \qquad \forall i \in \mathcal{I}. \tag{15}$$

We take $\alpha = 0.5$, $\gamma = 100$ and for all other parameters we take the same values as reported by Ben-Tal et al. (2013). This means that there are three scenarios per item, corresponding to low ($d_i^{(s)} = 4$), medium ($d_i^{(s)} = 8$) and high ($d_i^{(s)} = 10$) demand. The other parameters are listed in Table 1. We solve the problem for different values of ρ, varying between 0.000 and 0.030 in steps of 0.0001, where $\rho = 0$ corresponds to the nonrobust formulation where the estimates $\hat{p}_i^{(s)}$ are assumed to be exact, with AIMMS 3.11 and KNITRO 7.0. First, we check the conditions of Theorem 1. (COD-NV) is feasible for $\rho = 0$, and the solution is a Slater point of (COD-NV) for larger ρ. By (15) all w_{is} are forced to 0 if $x = 0$. Lastly, we have observed numerically that (COD-NV) is unbounded for ρ larger than 0.0306. This means that the primal problem is infeasible for ρ larger than 0.0306. The robust optimal $\boldsymbol{Q}$ and $\boldsymbol{u}$ are the elements of a KKT vector corresponding to constraints (13) and (14), respectively. As mentioned, we may have to verify the quality of the KKT vector. First, we check whether it is feasible to (R-NV). The constraint violation of (10) can be computed by maximizing a linear function over a convex set, and was found to be at most $1.5 \cdot 10^{-5}$ among all solutions. Second, we need to verify optimality by comparing the objective values of (R-NV) and (COD-NV). We observe both positive and negative relative differences of at most $2.0 \cdot 10^{-4}$. So, the quality of the KKT vector reported by the solver is accurate for this particular problem. The optimal order quantities and corresponding ordering costs are listed in Table 2 for different values of ρ.

For every solution we have uniformly sampled 10,000 p matrices from the uncertainty region, and computed the corresponding expected profit. Because implementing the robust solution requires a larger investment, the following comparison is based on the expected return, which is obtained by dividing the expected profit by the total ordering costs. The mean value and the range of these expected returns are listed in Table 3. As can be seen from this picture, the mean value for the nonrobust solution is often worse than the worst case for the robust solution. We will explain why the robust solution performs much better using the expected return of a single item. In the same way as for all items together, we have computed the expected return for item 3 (Figure 1). The peak at $\rho = 0.022$ is not a simulation inaccuracy, but it is caused by buying a larger number of items than for slightly larger or smaller ρ. The largest increase in expected return is between $\rho = 0$ and $\rho = 0.005$, for which the order quantity increases from 4.00 to 5.87 (Table 2). The profits for item 3 in the three scenarios are $(12, -8, -18)$ for $Q_3 = 4.00$, and $(3.6, 7.0, -3.0)$ for $Q_3 = 5.87$. So, with a slightly larger investment, the variation of the profit becomes much smaller, and hence, deviations in the probabilities on the scenarios have a smaller impact on the expected profit. This reduces the range of the expected return of the the robust solution, which can be seen in Figure 1. In total, six items show this behaviour. One of these (item 10) is more robust only for ρ between 0.003 and 0.012. For five items the robust order quantity is the same as the nonrobust order quantity and hence they do not have better robust performance. One item (item 2) has slightly worse robust performance.

Table 1: Parameter values for the multi-item newsvendor example

i	1	2	3	4	5	6	7	8	9	10	11	12
c_i	4	5	6	4	5	6	4	5	6	4	5	6
v_i	6	8	9	5	9	8	6	8	9	6.5	7	8
r_i	2	2.5	1.5	1.5	2.5	2	2.5	1.5	2	2	1.5	1
l_i	4	3	5	4	3.5	4.5	3.5	3	5	3.5	3	5
$\hat{p}_i^{(1)}$	0.375	0.250	0.375	0.127	0.958	0.158	0.485	0.142	0.679	0.392	0.171	0.046
$\hat{p}_i^{(2)}$	0.375	0.250	0.250	0.786	0.007	0.813	0.472	0.658	0.079	0.351	0.484	0.231
$\hat{p}_i^{(3)}$	0.250	0.500	0.375	0.087	0.035	0.029	0.043	0.200	0.242	0.257	0.345	0.723

Table 2: Optimal ordering cost and quantities for the multi-item newsvendor problem

ρ	cost	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}	Q_{11}	Q_{12}
0.000	391	8.00	8.00	4.00	8.00	4.00	8.00	4.00	8.00	4.00	8.00	7.03	8.00
0.005	412	8.00	8.00	5.87	8.00	4.00	8.00	5.69	8.00	4.00	7.01	8.00	8.34
0.010	421	8.00	8.00	6.20	8.00	4.00	8.00	6.12	8.00	4.00	7.55	8.00	8.85
0.015	430	8.00	8.00	6.39	8.00	4.00	8.00	6.36	8.00	4.00	8.00	8.00	9.62
0.020	440	8.00	8.00	7.10	8.00	4.00	8.00	7.31	8.00	4.00	8.00	8.00	10.00
0.025	453	8.00	8.00	7.36	8.00	4.00	8.00	8.00	8.00	5.51	8.00	8.00	10.00
0.030	469	8.00	9.49	8.00	8.00	4.00	8.00	8.00	8.00	6.26	8.00	8.00	10.00

Table 3: Simulation results of the expected return for the robust and nonrobust solutions for the multi-item newsvendor problem.

ρ	Robust solution			Nonrobust solution		
	min	mean	max	min	mean	max
0.000	0.2557	0.2557	0.2557	0.2557	0.2557	0.2557
0.005	0.2538	0.2695	0.2842	0.2369	0.2542	0.2737
0.010	0.2545	0.2753	0.2987	0.2269	0.2529	0.2795
0.015	0.2560	0.2806	0.3076	0.2169	0.2515	0.2855
0.020	0.2502	0.2855	0.3215	0.2121	0.2501	0.2903
0.025	0.2497	0.2815	0.3125	0.2088	0.2490	0.2955
0.030	0.2425	0.2809	0.3140	0.2026	0.2476	0.2900

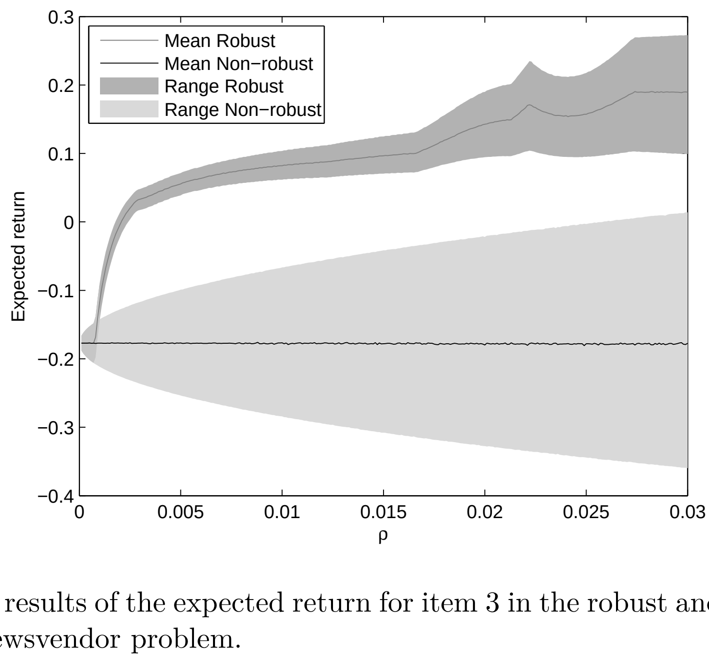

Figure 1: Simulation results of the expected return for item 3 in the robust and nonrobust solutions for the multi-item newsvendor problem.

Acknowledgments

Research of A. Ben-Tal is supported by the Israel–USA Science Foundation (BSF) Grant no. 2008302. The authors thank the anonymous referees and the associate editor for their useful comments.

References

A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *The Journal of Machine Learning Research*, 6:1705–1749, 2005.

A. Beck and A. Ben-Tal. Duality in robust optimization: primal worst equals dual best. *Operations Research Letters*, 37(1):1–6, 2009.

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

A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1):63–89, 2006.

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

A. Ben-Tal, D. den Hertog, A. M. B. de Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1–2):265–299, 2015.

D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

H. Blanc and D. den Hertog. On Markov chains with uncertain data. *CentER Discussion Paper*, 2008(050), 2008.

L. M. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

G. B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, 1963.

L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

J. E. Falk. Exact solutions of inexact linear programs. *Operations Research*, 24(4):783–787, 1976.

D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

R. N. Kaul, S. Kaur, and V. Lyall. Duality in inexact fractional programming with set-inclusive constraints. *Journal of Optimization Theory and Applications*, 50(2):279–288, 1986.

K. Matusita. On the notion of affinity of several distributions and some of its applications. *Annals of the Institute of Statistical Mathematics*, 19(1):181–192, 1967.

A. Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, volume I, pages 547–561. University of California Press, 1961.

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

M. Römer. Von der Komplexmethode zur robusten Optimierung und zurük. In *Beiträge zum Festkolloquium "Angewandte Optimierung" anlüslich des 65. Geburtstages von Prof. Dr. Rolf Rogge*, pages 37–61. Universität Halle–Wittenberg, 2010.

C. Singh. Convex programming with set-inclusive constraints and its applications to generalized linear and fractional programming. *Journal of Optimization Theory and Applications*, 38(1):33–42, 1982.

A. L. Soyster. A duality theory for convex programming with set-inclusive constraints. *Operations Research*, 22(4):892–898, 1974.

D. J. Thuente. Duality theory for generalized linear programs with computational methods. *Operations Research*, 28(4):1005–1011, 1980.