CentER

Discussion Paper

No. 2011-115

Robust counterparts of inequalities containing sums of maxima of linear functions

By Bram L. Gorissen, Dick den Hertog

October, 2011

TILBURG ◆ UNIVERSITY

Robust counterparts of inequalities containing sums of maxima of linear functions

Bram L. Gorissen,* Dick den Hertog

Tilburg University, Department of Econometrics and Operations Research, 5000 LE Tilburg, Netherlands
{b.l.gorissen,d.denhertog}@tilburguniversity.edu

October 25, 2011

Abstract

This paper adresses the robust counterparts of optimization problems containing sums of maxima of linear functions and proposes several reformulations. These problems include many practical problems, e.g. problems with sums of absolute values, and arise when taking the robust counterpart of a linear inequality that is affine in the decision variables, affine in a parameter with box uncertainty, and affine in a parameter with general uncertainty.

In the literature, often the reformulation that is exact when there is no uncertainty is used. However, in robust optimization this reformulation gives an inferior solution and provides a pessimistic view. We observe that in many papers this conservatism is not mentioned. Some papers have recognized this problem, but existing solutions are either too conservative or their performance for different uncertainty regions is not known, a comparison between them is not available, and they are restricted to specific problems. We provide techniques for general problems and compare them with numerical examples in inventory management, regression and brachytherapy. Based on these examples, we give tractable recommendations for reducing the conservatism.

keywords: robust optimization; sum of maxima of linear functions; biaffine uncertainty; robust conic quadratic constraints

JEL classification: C61

1 Introduction

Robust Optimization (RO) first appeared in Soyster (1973), and after receiving very little attention in the subsequent decades it has been an active research area since Ben-Tal and Nemirovski (1999) and El Ghaoui and Lebret (1997) started publishing new results in the late nineties. In its basic form, it requires a solution of an optimization problem to be feasible for any realization of the uncertain parameters in a given uncertainty set. For several choices of the uncertainty region this leads to tractable problems; for instance the robust counterpart (RC) of an LP with polyhedral uncertainty can be reformulated as an LP and the RC of an LP with ellipsoidal uncertainty is a conic quadratic program (CQP) (Ben-Tal et al. 2009a, p. 21).

In RO two distinct formulations that are equivalent in the nonrobust case may have different RCs. This is the case for optimization problems containing the sum of maxima of linear functions, of which the sum of absolute values is a special case ($|x| = \max\{x, -x\}$). These

---

*corresponding author

problems arise in inventory management, supply chain management, regression models, tumour treatment, and many other practical situations.

Because RO is applied constraint-wise w.l.o.g. and the objective function can always be formulated as a constraint (Ben-Tal et al. 2009a, p. 11), we focus on the following robust constraint:

$$\ell(\zeta, x) + \sum_{i \in I} \max_{j \in J} \{\ell_{ij}(\zeta, x)\} \leq d \qquad \forall \zeta \in \mathcal{Z}, \qquad (1)$$

where ℓ and ℓ_{ij} are biaffine functions in the uncertain parameter $\zeta \in \mathbb{R}^L$ and the decision variable $x \in \mathbb{R}^n$, $d \in \mathbb{R}$ is the right hand side, and $\mathcal{Z} \subset \mathbb{R}^L$ is a user-specified uncertainty region, e.g. a box or an ellipsoid. We can assume this uncertainty region to be closed and convex w.l.o.g., as the left hand side is convex in ζ and consequently the worst case is always located at an extreme point of the uncertainty region. In the remainder of this paper our objective is always to minimize d, but all methods can still be applied when a different objective is used.

In the literature, often the following RC is used, obtained from a reformulation with analysis variables y_i that is exact when there is no uncertainty:

$$\text{(RC-R)} \qquad \ell(\zeta, x) + \sum_{i \in I} y_i \leq d \qquad \forall \zeta \in \mathcal{Z} \qquad (2)$$

$$y_i \geq \ell_{ij}(\zeta, x) \qquad \forall \zeta \in \mathcal{Z} \quad \forall i \in I \quad \forall j \in J. \qquad (3)$$

In this formulation, to which we refer as the RC-R (RC reformulation), $y_i \in \mathbb{R}$ is a fixed variable, taking the worst case value of the ith term of the sum. In many cases the terms of the sum do not all reach their worst case value in the same realisation of the uncertain parameter ζ, and therefore the RC-R is a conservative reformulation of constraint (1), i.e. a solution (x, d) that is feasible for RC-R is also feasible for constraint (1), but not necessarily vice versa. However, this reformulation is oftenly used without mentioning its conservatism.

Bertsimas and Thiele (2006) use the RC-R for a robust inventory problem which includes the sum of holding and backlogging costs: $\sum_{i=1}^{n} \max\{c_h x_i, -c_b x_i\}$, where x_i is the inventory level at time period i and c_h and c_b are the holding and backlogging costs per time period respectively. Their uncertainty region is the intersection of $\{\zeta : ||\zeta_{1:i}||_1 \leq \Gamma_i\}$ $\forall i$ and $\{\zeta : ||\zeta||_\infty \leq 1\}$, where $\zeta_{1:i}$ is the vector consisting of the first i elements of ζ, and Γ_i is not necessarily integer (budgeted uncertainty). The reason they do not just put a bound on $||\zeta||_1$, is because they notice that the role of this bound disappears for constraints setting the cost for the first few time periods, because the few uncertain parameters included in those constraints can all take their boundary values. The restriction imposed by the $||\cdot||_\infty$ makes their model less robust against large deviations in early time periods. Moreover, their formulation allows that the values of different analysis variables take values based on different realizations of the uncertain parameter. The same uncertainty region is also used by Bertsimas and Thiele (2004) for a supply chain model, by Thiele (2004) for supply chains and revenue management, by Alem and Morabito (2012) for a production planning problem under uncertain demand, and by Wei et al. (2009) for a slightly more complicated inventory model in which items can be returned and remanufactured or disposed. With the exception of (Thiele 2004, p.37), none of these papers mention the conservatism of the formulation.

Ng et al. (2010) treat a lot allocation problem where each order is assigned to one or more production locations before the production capacity of each location is fully known. Every production location is assigned at most one order. Their uncertainty region is an ellipsoid over all production locations, while an analysis variable is used for every order, so their formulation is too conservative.

Kropat and Weber (2008) consider a robust linear cluster regression model using the sum of absolute values with polyhedral and ellipsoidal uncertainty regions. They introduce an analysis variable for every absolute value, which is too conservative because every uncertain parameter affects two absolute values.

Ben-Tal et al. (2005) have a multi-period inventory problem and use affine decision rules for the actual decisions (AARC). For every time period there are analysis variables indicating the

costs in that time period. These costs are given by the maximum of holding costs and shortage costs, which both are linear functions of past demand. In other words, each analysis variable is the maximum of two linear functions of the uncertain parameter. They note that an analysis variable should therefore be replaced with a function being the maximum of the two linear functions. Because that would lead to a very complicated robust counterpart, they replace the analysis variables with linear decision rules instead, which is conservative. Replacing analysis variables with linear decision rules was also done by Ben-Tal et al. (2009b) and Ben-Tal et al. (2011a).

Bienstock and Özbay (2008) were the first to identify and eliminate the conservatism of the RC-R. The idea behind their solution is that it suffices to make constraint (1) hold for just the vertices of the uncertainty region. The constraint then also holds for all other elements in the (convex) uncertainty region, because the constraint is convex in ζ. They generalize it to an iterative procedure that only adds a subset of the vertices, which they successfully tested for computing basestock levels under budgeted uncertainty. It is not yet known how well this iterative procedure performs on other problems or different uncertainty regions.

In this paper we identify a frequently existing conservatism in the formulation of a robust optimization problem. We provide an overview of exact (non-conservative) formulations, approximations, and iterative methods in Section 2. It is not possible to compare the objective value of different formulations in RO, so we introduce a new performance number which is independent of the reformulation in Section 3. The application scope of this paper is extended in Section 4, where we show that our methods can be applied to a conic quadratic constraint with box uncertainty. We evaluate the methods in Section 5 on some small toy problems and three larger problems. We give conclusions in Section 6, and show how the robust counterparts in the aforementioned papers could be improved.

The LP, MIP and CQP problems in this paper have been solved with AIMMS 3.11 FR3 x32 with ILOG CPLEX 12.1 unless stated otherwise. SDP problems have been modelled with CVX (Grant and Boyd (2010)) and solved with SDPT3 (YALMIP (Löfberg (2010)) would nowadays be a better choice as it allows the user to specify a linear constraint with quadratic uncertainty with an ellipsoidal uncertainty region directly). Reported computing times have been obtained under Windows XP SP3 on an Intel Core2 Duo E6400 (2.13 GHz) processor and 2 GB of RAM.

2 Solution approaches

2.1 Exact reformulations

2.1.1 Vertex enumeration

Vertex enumeration is an exact solution method introduced by Bienstock and Özbay (2008) that is powerful especially when the uncertainty region has a small number of vertices. Let $\zeta^s (s \in S)$ denote the vertices, and consider the following reformulation of constraint (1):

$$\text{(Vertex enumeration)} \qquad \ell(\zeta^s, x) + \sum_{i \in I} y_i^s \leq d \qquad \qquad \forall s \in S \qquad (4)$$

$$y_i^s \geq \ell_{ij}(\zeta^s, x) \qquad \forall i \in I \quad \forall j \in J \quad \forall s \in S.$$

This reformulation makes constraint (1) hold for all vertices of the uncertainty region. Constraint (1) is convex in ζ, so it also holds for all other points in the uncertainty region.

2.1.2 Enumeration of robust linear constraints

The RC-R is inexact because it has analysis variables that may take values corresponding to different worst case scenarios. A constraint with a single max function does not suffer from this problem, because an equivalent set of linear constraints can be formulated without analysis variables. An exact reformulation of RC (1) can be obtained by first rewriting it as a constraint

with a single $\max\{\cdot\}$ function by enumerating all combinations, and then applying RO to the reformulation:

$$\text{(EORLC)} \qquad \ell(\zeta, x) + \sum_{i \in I} \ell_{i,j(i)}(\zeta, x) \le d \qquad \forall j(i) \in J \qquad \forall i \in I \qquad \forall \zeta \in \mathcal{Z}.$$

We call this the enumeration of robust linear constraints (EORLC) formulation. For example, with $I = \{1, 2\}$ and $J = \{1, 2\}$, the EORLC formulation has the following constraints:

$$\begin{aligned}
\ell(\zeta, x) + \ell_{1,1}(\zeta, x) + \ell_{2,1}(\zeta, x) &\le d & \forall \zeta \in \mathcal{Z} \\
\ell(\zeta, x) + \ell_{1,1}(\zeta, x) + \ell_{2,2}(\zeta, x) &\le d & \forall \zeta \in \mathcal{Z} \\
\ell(\zeta, x) + \ell_{1,2}(\zeta, x) + \ell_{2,1}(\zeta, x) &\le d & \forall \zeta \in \mathcal{Z} \\
\ell(\zeta, x) + \ell_{1,2}(\zeta, x) + \ell_{2,2}(\zeta, x) &\le d & \forall \zeta \in \mathcal{Z}.
\end{aligned}$$

A similar formulation is also given by Bienstock and Özbay (2008), where it was neglected for its exponential size. While the number of constraints $|J|^{|I|}$ indeed grows exponentially with the number of terms in the summation, it is effective for small $|I|$. There are situations in which I is small indeed, for instance with a planning horizon of up to ten periods. An example in the literature with small $|I|$ is the robust regression model by Kropat and Weber (2008). For this regression model, the exact method increases the number of constraints with only a factor four regardless of the number of observations, while reducing the number of analysis variables with a factor two, compared to the pessimistic RC-R applied in the paper.

EORLC has a strong relation to vertex enumeration. In fact, it is exactly vertex enumeration but on a different set. Constraint (1) can be formulated as:

$$\ell(\zeta, x) + \sum_{i \in I} \sum_{j \in J} \lambda_{ij} \ell_{ij}(\zeta, x) \le d \qquad \forall \lambda_i \in \Delta^{|J|-1} \quad \forall i \in I \quad \forall \zeta \in \mathcal{Z},$$

where $\Delta^{|J|-1} = \{\lambda_i \in \mathbb{R}^{|J|} : \sum_{j \in J} \lambda_{ij} = 1, \lambda_{ij} \ge 0\}$ (the standard simplex in $\mathbb{R}^{|J|}$). The vertices of the simplex are given by unit vectors. If λ_i is a unit vector, $\sum_{j \in J} \lambda_{ij} \ell_{ij}(\zeta, x)$ simplifies to a single $\ell_{ij}(\zeta, x)$. It follows that vertex enumeration on the $|J|$ simplices gives the EORLC reformulation.

There is one special remark for inequalities of the form $a \max\{0, \ell(\zeta, x)\} + b \max\{0, -\ell(\zeta, x)\} \le d$ with $a, b > 0$, which are often used for holding and backlogging costs. Before enumerating robust linear constraints, the inequality should be reformulated as $\max\{a\ell(\zeta, x), -b\ell(\zeta, x)\} \le d$ in order to reduce the number of constraints.

2.1.3 Cases with special structure

There are several special cases of (1) that allow an exact reformulation. This section lists a few general cases. More specific cases can be found in (Ben-Tal et al. 2009a, Ch. 12.2) and in Xu et al. (2009).

The first case is when the uncertainty region $\mathcal{Z}$ is the direct product of sets $\mathcal{Z}_i$, where ℓ_{ij} only depends on ζ_i in $\mathcal{Z}_i$ and x. The RC-R is exact because all analysis variables can take their worst case values simultaneously.

The second case is when the inequality contains the sum of absolute values of linear functions of ζ_i and x:

$$\ell(\zeta, x) + \sum_{i \in I} |\alpha_i(x) + \beta_i(x)\zeta_i| \le d \qquad \forall \zeta \in \mathcal{Z}, \tag{5}$$

where $\alpha_i : \mathbb{R}^n \to \mathbb{R}$ and $\beta_i : \mathbb{R}^n \to \mathbb{R}$ are linear functions, and the uncertainty region $\mathcal{Z}$ is closed under changing the sign of one or more vector elements (a box and an ellipsoid are examples of such sets). We will call $\mathcal{Z}$ symmetric around $\zeta = 0$. Note that the assumption that the symmetry is around 0 is made w.l.o.g. The following constraint is equivalent to (5):

$$\ell(\zeta, x) + \sum_{i \in I} \{|\alpha_i(x)| + \beta_i(x)\zeta_i\} \le d \qquad \forall \zeta \in \mathcal{Z}.$$

4

Equivalence is readily checked by conditioning on the sign of $\alpha_i(x)$. The formulation where absolute values are replaced with analysis variables is equivalent, since the analysis variables do not depend on ζ.

The third case is when the sum of maxima of linear functions is the sum of a common nonnegative linear function of ζ and the product of a common nonnegative linear function of ζ and a linear function of x:

$$\ell(\zeta, x) + \sum_{i \in I} \max_{j \in J}\{\alpha_i(\zeta) + \beta_i(\zeta)\ell_{ij}(x)\} \leq d \qquad \forall \zeta \in \mathcal{Z},$$

where $\alpha_i : \mathbb{R}^L \to \mathbb{R}$ and $\beta_i : \mathbb{R}^L \to \mathbb{R}_+$ are linear functions. The common functions of ζ can be placed outside the $\max\{\cdot\}$ expression:

$$\ell(\zeta, x) + \sum_{i \in I} \left[\alpha_i(\zeta) + \beta_i(\zeta)\max_{j \in J}\{\ell_{ij}(x)\}\right] \leq d \qquad \forall \zeta \in \mathcal{Z},$$

and the RC-R of this constraint is exact. If the range of β_i is $\mathbb{R}$ instead of $\mathbb{R}_+$, then $\max_{j \in J}\{\ell_{ij}(x)\}$ should be $\min_{j \in J}\{\ell_{ij}(x)\}$ when $\beta_i(\zeta) < 0$. This can be modelled as follows. Given a subset $I_+ \subseteq I$, we define a set which consists of those ζ for which $\beta_i(\zeta) \geq 0$ for $i \in I_+$, and $\beta_i(\zeta) \leq 0$ for $i \in I \backslash I_+$:

$$\mathcal{Z}(I_+) = \mathcal{Z} \cap \{\zeta : \beta_i(\zeta) \geq 0 \quad \forall i \in I_+, \quad \beta_i(\zeta) \leq 0 \quad \forall i \in I \backslash I_+\}.$$

Note that

$$\mathcal{Z} = \bigcup_{I_+ \subseteq I} \mathcal{Z}(I_+).$$

The constraint can now be written as:

$$\ell(\zeta, x) + \sum_{i \in I}\alpha_i(\zeta) + \sum_{i \in I_+}\beta_i(\zeta)\max_{j \in J}\{\ell_{ij}(x)\} + \sum_{i \in I \backslash I_+}\beta_i(\zeta)\min_{j \in J}\{\ell_{ij}(x)\} \leq d \qquad \forall \zeta \in \mathcal{Z}(I_+) \qquad \forall I_+ \subseteq I.$$

The number of constraints is $2^{|I|}$ (one for each $I_+ \subseteq I$), which is less than the $|J|^{|I|}$ constraints obtained with EORLC. The max and min expressions do not depend on ζ, so the reformulation with analysis variables is exact. Each constraint is still convex despite the min expressions, because their coefficients $\beta_i(\zeta)$ are negative.

2.2 Approximations

The variables y_i in constraints (2)-(3) have been introduced for modeling the $\max\{\cdot\}$ expression. We do not need to know their values because they do not correspond to a "here and now" decision. Only x has to be known for implementing a solution. The values of y_i may be adjusted according to the realisation of the uncertain parameter ζ as long as the constraints hold for every realisation of ζ in the perturbation set (Adjustable RC). As first proposed by Ben-Tal et al. (2005) for a multi-stage problem and later also done by Ben-Tal et al. (2009b) and Ben-Tal et al. (2011a), we can make y_i an affine function of ζ, which leads to the Affinely Adjustable RC Reformulation (AARC-R). After substituting $y_i = v_i + w_i^{\intercal}\zeta$ (with decision variables $v_i \in \mathbb{R}$ and $w_i \in \mathbb{R}^L$), constraints (2)-(3) become:

$$\text{(AARC-R)} \qquad \ell(\zeta, x) + \sum_{i \in I}\left(v_i + w_i^{\intercal}\zeta\right) \leq d \qquad \qquad \forall \zeta \in \mathcal{Z}$$

$$v_i + w_i^{\intercal}\zeta \geq \ell_{ij}(\zeta, x) \qquad \forall \zeta \in \mathcal{Z} \quad \forall i \in I \quad \forall j \in J.$$

This substitution gives a less conservative reformulation, while the robust counterpart is in the same class of problems as the RC-R albeit with a larger number of variables. If ℓ_{ij} does not depend on one or more components of ζ for all j for some fixed i, the computational complexity

can seemingly be reduced by making y_i a function of only those components that directly affect it. However, it is easy to construct an example where this reduction introduces more conservatism if the components are dependent.

There are two different approaches that lead to the same result. The first approach is considering the Fenchel dual problem of maximizing the left hand side of constraint (1) over ζ in $\mathcal{Z}$. The resulting constraint is conservative because a non-concave maximization problem is dualized. We give the full derivation and a proof of equivalence to the AARC-R in Appendix A. The other approach is derived in Appendix B.

Ben-Tal et al. (2005) not only replace analysis variables with linear decision rules, but also replace the actual decision variables since they solve a multistage problem. They evaluate the quality of their formulation by also computing the optimal (possibly nonaffine) policy and corresponding worst case costs. These costs are the same as the value of the AARC for almost all of their data sets, from which they conclude that an affine policy provides excellent results. This conclusion is only about the quality of the decision rules for the actual decision variables, and not about the decision rules for the analysis variables. We draw the additional conclusion that the AARC-R is exact for almost all cases in their data set.

An approach that is even less conservative than an affine decision rule, is a quadratic decision rule:

$$y_i = v_i + w_i^\intercal \zeta + \zeta^\intercal W_i \zeta,$$

where $v_i \in \mathbb{R}$, $w_i \in \mathbb{R}^L$, and $W_i \in \mathbb{R}^{L \times L}$ are new analysis variables. This is called a Quadratically Adjustable RC reformulation (QARC-R) which is known to be tractable for ellipsoidal uncertainty and (under some restrictions) for box uncertainty. When the uncertainty set is ellipsoidal, the RC is an SDP with $|I||J| + 1$ variables of size $(L + 1) \times (L + 1)$ (Ben-Tal et al. 2009a, p. 382). For box uncertainty, when the quadratic terms are restricted such that each element of ζ is multiplied with itself and at most one other element of ζ, we can use the result by Yanıkoğlu et al. (2011) to write the RC as an SDP with $(|I||J| + 1)\lceil L/2 \rceil$ variable matrices of size three.

Another way to reduce the conservatism of the RC-R is by first combining several max expressions before reformulating. In order to do this, partition I into $|G|$ groups: $I = \bigcup_{g \in G} I_g$, where the groups are mutually disjoint: $I_{g_1} \cap I_{g_2} = \emptyset$ for $g_1, g_2 \in G, g_1 \neq g_2$, and partition in such a way that $|I_g|$ is small for all g. Now introduce analysis variables for all g:

$$\ell(\zeta, x) + \sum_{g \in G} y_g \leq d \qquad\qquad \forall \zeta \in \mathcal{Z}$$

$$y_g \geq \sum_{i \in I_g} \max\{\ell_{ij}(\zeta, x)\} \qquad\qquad \forall \zeta \in \mathcal{Z} \quad \forall g \in G.$$

Because the cardinality of I_g is small, the sum of max expression within each constraint can be transformed into a single max (EORLC). Each constraint in this reformulation is therefore tractable and can be solved exactly, but the conservatism comes from the analysis variables y_g. These analysis variables can also be written as a linear, quadratic or more general function of ζ.

2.3 Iterative methods

In this section we describe two iterative methods. The first one is based on vertex enumeration, the other one on EORLC.

An issue with box uncertainty is that the number of vertices grows exponentially in the time horizon. Also for budgeted uncertainty, which is described in Section 1, the number of vertices quickly becomes very large. Bienstock and Özbay (2008) proposed an iterative alternative for vertex enumeration that only works with a subset of the vertices. We give a general description of their procedure in Algorithm 1.

Algorithm 1 Iterative vertex enumeration

Require: A linear program LP with constraints (4)
 1: $S := \{\zeta^{nom}\}$ (the nominal value)
 2: $k := 0$
 3: **repeat**
 4:    $k := k + 1$
 5:    Solve LP
 6:    Let x^* be the minimizer of LP, and LB be the value of d in LP
 7:    Solve $MIP := \max_{\zeta \in \mathcal{Z}} \left\{ \ell(\zeta, x^*) + \sum_{i \in I} \max_{j \in J} \{\ell_{ij}(\zeta, x^*)\} \right\}$
 8:    Let ζ^k be the maximizer, and UB be the value of MIP
 9:    $S := S \cup \{\zeta^k\}$
10: **until** $UB - LB < \varepsilon$

While the algorithm is running, LB is a lower bound on the optimal value of d in (4) because it is the value of a relaxation of the original problem, and UB is an upper bound on the optimal d, because it is the maximum for some x and not for the optimal x. The difference $UB - LB$ indicates the current violation of the constraint. If ε is set to a larger value, the algorithm terminates quicker but does not give the optimal solution. If $\varepsilon = 0$ and the algorithm terminates, the final solution is robust feasible and robust optimal. It is possible that this algorithm still enumerates all constraints, but we have not encountered problems for which this is the case. Bienstock and Özbay (2008) find that for their problem the number of iterations does not increase when the time horizon is increased, and is around four on average even for 150 time periods. The number of vertices of their uncertainty region can not be deducted from their report. In our numerical examples we often find a larger number of iterations. We argue that the specific structure of their problem and their specific uncertainty region may explain the effectiveness of their method, and it is not clear if the method works equally well on other problems. The problem they solve is an inventory problem with known, time dependent ordering cost, holding cost and backlogging cost, where they try to determine the optimal time invariant order-up-to level while demand is uncertain. An order-up-to-policy is computationally-wise simple, because the inventory level at the beginning of each period is the same. Given the order-up-to-level, for every time period the cost of the use of one unit of budget can be computed. In the worst case, the budget is used at those time periods where the cost is maximal. For small changes in the order-up-to-level, the costs of one unit of budget do not vary much. Hence, the set of time periods where the budget is used, can be estimated accurately in the first one or two iterations. It is slightly more complex if the initial inventory is much higher than the order-up-to-level, because in that case the starting inventory for the first time periods does not equal the order-up-to-level. We can then split the time horizon into three periods: an initial period in which the inventory level is above the order-up-to-level, then an uncertain period in which it depends on the demand in the past whether the inventory level falls below the order-up-to-level, and then the final period in which the initial inventory level is known for every time period. The uncertainty mostly affects the second period, which explains why the method adds the same number of demand trajectories irrespective of the time horizon. They also consider the uncertainty region with interval demand and a possibility of peak demand of known size at unknown periods each at least a defined number of periods apart. The solution with this uncertainty region is again simple: demand in the first few periods should be as small as possible in order to maximize holding costs, until a point where maximal demand (or peak demand) incurs more backlogging costs than holding costs with small demand. From that moment onwards, the inventory level at the beginning of each period is known, and the costs of small demand, large demand and peak demand can be computed for every time period. The same iterative method was used by Bohle et al. (2010), again for budgeted uncertainty, but solution times and the number of generated constraints are not reported. It is still an open question how well this method works on other problems, and for non polyhedral uncertainty regions.

The most time consuming step is solving MIP, since this involves the maximization of a convex function. As pointed out by Bienstock and Özbay (2008), for Algorithm 1 it is not necessary to find the optimal solution. It suffices to find any ζ for which the maximization problem has a larger objective value than LB, because it corresponds to a violated constraint in the relaxation, and adding that constraint strengthens the relaxation. Bienstock et al. report that also the problem LP does not need to be solved to optimality. Optimization can stop as soon as the objective value is less than UB, because at that point the current solution is already better than the solution in the previous run.

The second iterative approach is EORLC. If $|I|$ becomes too large, an iterative method can add only a subset of the robust linear constraints. It starts with the nominal problem and adds new robust constraints until the solution does not violate any of the constraints of the full enumeration method. We give a general description of our procedure in Algorithm 2.

Algorithm 2 Adding robust linear constraints

Require: A linear program LP with constraints (4)
 1: $S := \{\zeta^{nom}\}$ (the nominal value)
 2: $k := 0$
 3: **repeat**
 4:     $k := k + 1$
 5:     Solve LP
 6:     Let x be the minimizer of LP, and LB be the value of d in LP
 7:     Solve $MIP := \max_{\zeta \in \mathcal{Z}} \left\{ \ell(\zeta, x) + \sum_{i \in I} \max_{j \in J} \{ \ell_{ij}(\zeta, x) \} \right\}$
 8:     Let ζ^k be the maximizer, and UB be the value of MIP
 9:     Let j_i^k be the maximizer of $\max_{j \in J} \ell_{ij}(\zeta^k, x)$
10:     Add the following robust constraint to LP:

$$\sum_{i \in I} \ell_{i, j_i^k}(\zeta, x) \leq d \qquad \forall \zeta \in \mathcal{Z}$$

11: **until** $UB - LB < \varepsilon$

Similar to Algorithm 1, a lower and upper bound are reported and the stopping criterion can be adjusted. Also for this algorithm, we have not encountered numerical examples in which all constraints are enumerated. The main difference between this algorithm and Algorithm 1 is the constraint that is added in every iteration. The algorithms can easily be combined so that two sets of constraints are added per iteration.

Again solving MIP is the most time consuming step, and also here it does not need to be solved to optimality as long as the value is larger than LB. Also LP does not need to be solved to optimality.

3   The true robust value

This section describes a value that can be used to compare different solutions. For a minimization problem containing sums of maxima of linear functions, we define the true robust value of a solution to be the maximum over ζ in $\mathcal{Z}$ of the unreformulated problem with all decision variables fixed. We have assumed that in this paper our objective is always to minimize d, an analysis variable at the right hand side of (1), so in our case the true robust value of a solution x is:

$$v_{true}(x) = \max_{\zeta \in \mathcal{Z}} \left\{ \ell(\zeta, x) + \sum_{i \in I} \max_{j \in J} \{ \ell_{ij}(\zeta, x) \} \right\}. \tag{6}$$

8

This value and its maximizer appear in Algorithm 1 and 2. Determining this value is a difficult problem, because it requires the maximization of a convex function over a convex set. The global maximum over $\zeta \in \mathcal{Z}$ of the sum of maxima of linear functions does not necessarily coincide with a maximum over $\zeta \in \mathcal{Z}$ of one of those linear functions. One way to obtain the exact value is by solving the following optimization problem with integer variables:

$$
\begin{aligned}
\max \quad & \ell(\zeta, x) + \sum_{i \in I} y_i \\
\text{s.t.} \quad & y_i \leq \ell_{ij}(\zeta, x) + M(1 - z_{ij}) && \forall i \in I \quad \forall j \in J \\
& \sum_{j \in J} z_{ij} = 1 && \forall i \in I \\
& x \in \mathbb{R}^n, \zeta \in \mathcal{Z}, y \in \mathbb{R}^{|I|}, z_{ij} \in \{0,1\}^{|I| \times |J|},
\end{aligned}
$$

which is a MIP for polyhedral uncertainty, and a MIQCP for ellipsoidal uncertainty. By using CPLEX, the worst case ζ can be determined in less than a minute for $|I| = 50$ and $|J| = 3$ for box uncertainty of dimension 50, and in less than a minute for $|I| = 20$ and $|J| = 2$ for ellipsoidal uncertainty of dimension 20.

Another way of obtaining the exact value is by considering $|J|^{|I|}$ linear optimization problems, e.g. for $I = \{1, 2\}$ and $J = \{1, 2\}$ consider the following four optimization problems:

$$
\begin{aligned}
& \max_{\zeta \in \mathcal{Z}} \ell(\zeta, x) + \ell_{1,1}(\zeta, x) + \ell_{2,1}(\zeta, x), \\
& \max_{\zeta \in \mathcal{Z}} \ell(\zeta, x) + \ell_{1,1}(\zeta, x) + \ell_{2,2}(\zeta, x), \\
& \max_{\zeta \in \mathcal{Z}} \ell(\zeta, x) + \ell_{1,2}(\zeta, x) + \ell_{2,1}(\zeta, x), \text{ and} \\
& \max_{\zeta \in \mathcal{Z}} \ell(\zeta, x) + \ell_{1,2}(\zeta, x) + \ell_{2,2}(\zeta, x).
\end{aligned}
$$

Each of these problems is easy, because the maximum of an affine function over a box or an ellipsoid can be computed in a few operations. The true robust value is the largest value of the computed maxima. To get an idea of the performance of this method, we have created a single threaded C++ application that randomly selects coefficients from the interval $[-100, 100]$ of an affine function $f : \mathbb{R}^{20} \to \mathbb{R}$, maximizes that function, randomly selects new coefficients, et cetera. Only the running time of the maximization step is measured. It can maximize 2×10^8 affine functions per minute for box uncertainty, and 5×10^7 affine functions per minute for ellipsoidal uncertainty. Comparing this speed to the speed of the optimization problem with integer variables, the latter is 10^{15} times faster for box uncertainty when $|I| = 50$ and $|J| = 3$, but 50 times slower for ellipsoidal uncertainty when $|I| = 20$ and $|J| = 2$. After changing the solver settings such that the MIQCP is solved by checking all leaves of the branch&bound tree, both approaches are solved at comparable speed.

If determining $v_{true}(x)$ is intractable, bounds can still be obtained. Filling in any ζ from the set $\mathcal{Z}$ at the right hand side of equation (6), for instance the nominal value, gives a lower bound. If the dimension of ζ is small, the bound can be improved with global optimization techniques. Upper bounds can be obtained by fixing x in any conservative reformulation of constraint (1), like the RC-R, AARC-R or QARC-R, and optimizing for the analysis variables.

In the remainder of this paper, v_{true} and x_{true}, v_{nom} and x_{nom}, v_{RC-R} and x_{RC-R}, v_{AARC-R} and x_{AARC-R}, v_{QARC-R} and x_{QARC-R}, and v_I and x_I denote the value and the solution of the true optimization problem, the nominal optimization problem, the RC-R, the AARC-R, the QARC-R and an iterative procedure, respectively.

4 RC of a linear constraint with biaffine uncertainty

Constraint (1) may appear in RO itself when a constraint has biaffine uncertainty, and the uncertainty region of one parameter is a box. In general, such a constraint can be written as:

$$\tilde{\ell}(\zeta^{(1)}, \zeta^{(2)}, x) \leq d \qquad \forall \zeta^{(1)} : \left\|\zeta^{(1)}\right\|_{\infty} \leq \rho \quad \forall \zeta^{(2)} \in \mathcal{Z}, \tag{7}$$

where $\tilde{\ell} : \mathbb{R}^{|I|} \times \mathbb{R}^{L} \times \mathbb{R}^{n} \to \mathbb{R}$ is a triaffine function, ρ is the radius of the box, and $\mathcal{Z}$ is the uncertainty region of $\zeta^{(2)}$. Constraints with biaffine uncertainty have never been investigated before to the best of our knowledge. To show the equivalence to constraint (1), choose ℓ, ℓ_i and I in such a way that constraint (7) is equivalent to:

$$\ell(\zeta^{(2)}, x) + \sum_{i \in I} \zeta_i^{(1)} \ell_i(\zeta^{(2)}, x) \leq d \qquad \forall \zeta^{(1)} : \left\|\zeta^{(1)}\right\|_{\infty} \leq \rho \quad \forall \zeta^{(2)} \in \mathcal{Z},$$

and maximize the left hand side with respect to $\zeta^{(1)}$:

$$\ell(\zeta^{(2)}, x) + \sum_{i \in I} \rho|\ell_i(\zeta^{(2)}, x)| \leq d \qquad \forall \zeta^{(2)} \in \mathcal{Z}, \tag{8}$$

which is indeed in the form of constraint (1). This RC includes many practical problems, of which we give four examples.

The first example is a constraint $a^{\mathsf{T}} x \leq d$ where each element of a has been estimated using a regression model with the same regressors ζ for every element: $a_i = \beta_i^0 + (\beta^{(i)})^{\mathsf{T}} \zeta + \varepsilon_i$ ($\beta_i^0 \in \mathbb{R}, \beta^{(i)} \in \mathbb{R}^L, \varepsilon_i \in \mathbb{R}$), in that case the constraint is:

$$(\beta^0 + B\zeta + \varepsilon)^{\mathsf{T}} x \leq d.$$

We assume that the coefficients $\beta^0 \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n \times L}$, the matrix having $\beta^{(i)}$ as rows, have been estimated with some error, and also that the regressors ζ are not fully known. Current literature on robust regression only focusses on estimating the regression parameters, and not on using an estimate in a constraint. The constraint can be written as constraint (7) if β^0, B, ζ and ε lie in some specified general uncertainty region, with box uncertainty on either B or ζ.

The second example appears in Adjustable RCs. Consider a robust constraint

$$\ell(\zeta^{(1)}, \zeta^{(2)}, x) + b^{\mathsf{T}} y \leq d \qquad \forall \zeta^{(1)} : \left\|\zeta^{(1)}\right\|_{\infty} \leq \rho \quad \forall \zeta^{(2)} \in \mathcal{Z},$$

where $y \in \mathbb{R}^{m_1}$ is an adjustable variable that represents a "wait and see" decision that can be made after $\zeta^{(1)}$ and $\zeta^{(2)}$ are (partially) observed, and b is a vector of known coefficients (fixed recourse). Determining the true optimal policy for y as a function of $\zeta^{(1)}$ and $\zeta^{(2)}$ is often intractable, which is why a suboptimal y is often determined by a parameterization, i.e. y is written as a function of which the coefficients are decision variables. The first parameterization in RO was proposed by Ben-Tal et al. (2004) who proposed an affine decision rule, which results in a problem which is in the same class (LP, CQP or SDP, depending on the uncertainty region) as the problem with y as a "here and now" decision variable. This was later extended to a quadratic decision rule with an ellipsoidal uncertainty region, resulting in an SDP (Ben-Tal et al. (2009a)), and to a polynomial decision rule of arbitrarily large degree restricted to uncertainty regions described by polynomial inequalities, resulting in a conservative SDP (Bertsimas et al. (2009)). The latter includes the biaffine decision rule $y = \ell'(\zeta^{(1)}, \zeta^{(2)}, v)$, where ℓ' is a function $\mathbb{R}^{|I|} \times \mathbb{R}^{L} \times \mathbb{R}^{m_2} \to \mathbb{R}^{m_1}$, and v is a vector of coefficients to be determined by the model. This decision rule could be very useful if a problem is affected by several sources of uncertainty. The advantage over affine decision rules is that the former includes cross terms of $\zeta^{(1)}$ and $\zeta^{(2)}$. Applying the results of Bertsimas et al. (2009) gives an SDP which is not only conservative, but also has a potentially large instance size. E.g. if $|I|$ (the dimension of $\zeta^{(1)}$) is very small but $\zeta^{(2)}$ is in a box of (large) dimension L, the conservative SDP has at least one variable matrix

of size $(|I|+1)L+|I|+1$ and over $2|L|$ smaller matrices, while our result is exact and gives a practically solvable LP. If instead $\zeta^{(2)}$ is in an ellipsoid of (large) dimension L, the SDP still has the large variable matrix of size $(|I|+1)L+|I|+1$ while our exact result gives a CQP for which efficient solvers are available.

The third example is a constraint with unknown coefficients and implementation error. Consider the following constraint:

$$\ell(\zeta^{(1)})^\intercal x \leq d \qquad \forall \zeta^{(1)} \in \mathcal{Z}_1, \tag{9}$$

where ℓ is a vector of n functions that are linear in the uncertain parameter $\zeta^{(1)}$. Now suppose that there is implementation error in x, i.e. instead of x we implement a vector of which component i is given by $x_i + \zeta_i^{(2)}$ (additive implementation error) or by $\zeta_i^{(2)} x_i$ (multiplicative implementation error). After substituting x, constraint (9) becomes:

$$\ell(\zeta^{(1)})^\intercal (x + \zeta^{(2)}) \leq d \qquad \forall \zeta^{(1)} \in \mathcal{Z}_1 \quad \forall \zeta^{(2)} \in \mathcal{Z}_2,$$

in case of additive implementation error, and

$$\sum_{i=1}^n \ell_i(\zeta^{(1)})\zeta_i^{(2)} x_i \leq d \qquad \forall \zeta^{(1)} \in \mathcal{Z}_1 \quad \forall \zeta^{(2)} \in \mathcal{Z}_2,$$

in case of multiplicative implementation error. Both constraints are special cases of constraint (7) if either $\mathcal{Z}_1$ or $\mathcal{Z}_2$ is a box.

The fourth example is the following robust constraint with box uncertainty:

$$\ell(\zeta^{(1)}, x) + \sum_{k \in K} \left|\left|\ell_k(\zeta^{(1)}, x)\right|\right|_l \leq d \qquad \forall \zeta^{(1)} : \left|\left|\zeta^{(1)}\right|\right|_\infty \leq \rho, \tag{10}$$

where ℓ_k is a vector of L linear functions, and l equals 1, 2 or ∞. In order to see that this constraint is equivalent to constraint (7), note that constraint (10) is a reformulation of the following constraint:

$$\ell(\zeta^{(1)}, x) + \sum_{k \in K} (\zeta_k^{(2)})^\intercal \ell_k(\zeta^{(1)}, x) \leq d \qquad \forall \zeta^{(1)} : \left|\left|\zeta^{(1)}\right|\right|_\infty \leq \rho \qquad \forall \zeta^{(2)} : \left|\left|\zeta_k^{(2)}\right|\right|_l^* \leq 1 \quad \forall k \in K,$$

where $||\cdot||_l^*$ is the dual norm, and $\zeta_k^{(2)}$ is a vector in $\mathbb{R}^L$ for all k in K. We will show how the results in this paper can be used for solving the robust constraint (10) for different choices of l.

For $l = 2$ and $|K| = 1$, constraint (10) is a robust conic quadratic constraint, for which the RC is known only in special cases, one of which is vertex enumeration (Ben-Tal et al. 2009a, p. 159). The case $|K| > 1$ has not been covered yet. A (conservative) reformulation with analysis variables for every $\left|\left|\ell_k(\zeta^{(1)}, x)\right|\right|_2$ reduces the problem to a problem with one linear constraint and $|K|$ robust conic quadratic constraints, each of which is of the form (10) with $|K| = 1$, so it can be reformulated using vertex enumeration. A different approach, that is not only exact but also results in a smaller problem than obtained by using analysis variables, is to apply vertex enumeration to constraint (10) directly. Vertex enumeration is exact because the constraint is convex in $\zeta^{(1)}$. If the dimension of the box is very large, (iterative) vertex enumeration is no longer tractable, so even the case $|K| = 1$ becomes unsolvable, and tractable conservative reformulations are not known. Our reformulation (8) allows the use of all approaches from Section 2.

For $l = 1$ constraint (10) is a special case of constraint (1), and we know from Section 2 how to solve it or how to find a conservative reformulation. The reformulation may be useful if it is solved with the RC-R, the AARC-R or the QARC-R, because the resulting formulation is very different. Vertex enumeration and EORLC are not useful on the reformulation (8), because they correspond with EORLC and vertex enumeration on the original constraint, respectively.

For $l = \infty$ constraint (10) is a special case of constraint (1) with $I = K$ and $J = \{1, 2, .., L\}$, and we know from Section 2 how to solve it or how to find a conservative reformulation. The

reformulation (8) may be useful if it solved with the RC-R, the AARC-R or the QARC-R, because the resulting formulation is very different. The reformulation allows to do vertex enumeration on $\zeta^{(2)}$, which may be faster than vertex enumeration in the original constraint (which is done on the vertices of $\zeta^{(1)}$) if L and $|K|$ are small relative to $|I|$.

For all three cases of l, it holds that when EORLC is used on the reformulation (8), the resulting constraints are the same ones resulting from vertex enumeration on constraint (10). This implies that reformulating (8) is a redundant step if it is followed by EORLC.

5 Numerical examples

5.1 Illustrative small problems

Consider the following toy optimization problem:

$$
\begin{aligned}
\text{(TOY1)} \quad &\min && d \\
&\text{s.t.} && d \geq \max\{x, x+\zeta\} + \max\{x, x-\zeta\} && \forall \zeta \in [-1,1] \\
&&& d \in \mathbb{R}, x \in \mathbb{R}_+.
\end{aligned}
$$

The optimal value of this problem is 1 ($x = 0, d = 1$). If we model this problem as an LP and then apply RO, we get the following model:

$$
\begin{aligned}
\text{(RC-R)} \quad &\min && y_1 + y_2 \\
&\text{s.t.} && y_1 \geq x && \forall \zeta \in [-1,1] \\
&&& y_1 \geq x + \zeta && \forall \zeta \in [-1,1] \\
&&& y_2 \geq x && \forall \zeta \in [-1,1] \\
&&& y_2 \geq x - \zeta && \forall \zeta \in [-1,1] \\
&&& y_1, y_2 \in \mathbb{R}, x \in \mathbb{R}_+.
\end{aligned}
$$

The optimal value of this model is 2 ($x = 0, y_1 = y_2 = 1$), so the RC-R is not exact. If we substitute $y_i = v_i + w_i \zeta$, the model becomes:

$$
\begin{aligned}
\text{(AARC-R)} \quad &\min && y \\
&\text{s.t.} && y \geq v_1 + w_1 \zeta + v_2 + w_2 \zeta && \forall \zeta \in [-1,1] \\
&&& v_1 + w_1 \zeta \geq x && \forall \zeta \in [-1,1] \\
&&& v_1 + w_1 \zeta \geq x + \zeta && \forall \zeta \in [-1,1] \\
&&& v_2 + w_2 \zeta \geq x && \forall \zeta \in [-1,1] \\
&&& v_2 + w_2 \zeta \geq x - \zeta && \forall \zeta \in [-1,1] \\
&&& v_1, v_2, w_1, w_2, y \in \mathbb{R}, x \in \mathbb{R}_+.
\end{aligned}
$$

The optimal value of this problem is 1 ($v_1 = v_2 = \frac{1}{2}, w_1 = \frac{1}{2}, w_2 = -\frac{1}{2}, y = 1$), which is the same as the optimal value of the original problem. So, in this case the AARC-R closes the gap. This is not always the case as the following example shows:

$$
\begin{aligned}
\text{(TOY 2)} \quad &\min && d \\
&\text{s.t.} && d \geq \max\{x, x+\zeta_1+\zeta_2\} + \max\{x, x+\zeta_1-\zeta_2\} \\
&&& \quad + \max\{x, x-\zeta_1+\zeta_2\} + \max\{x, x-\zeta_1-\zeta_2\} && \forall \zeta \in [-1,1]^2 \\
&&& d \in \mathbb{R}, x \in \mathbb{R}_+.
\end{aligned}
$$

We obtain $v_{true} = 2$, $v_{RC-R} = 8$ and $v_{AARC-R} = 4$. The AARC-R is an improvement over the RC-R, but is not exact. The QARC-R, which can be derived by reparameterizing the uncertainty region to $[0,1]^2$, applying the work of Yanıkoğlu et al. (2011) to obtain an LMI description of the uncertainty region, and formulating the RC by dualization as demonstrated in Appendix C, gives $v_{QARC-R} = 2.83$. While this value is much closer to 2 than the value of the AARC-R, it is still inexact.

5.2 Least absolute deviations regression with errors-in-variables

An errors-in-variables regression model is a model $y_i = \beta_0 + \beta_1 x_i^* + \varepsilon_i$ ($\varepsilon_i \sim N(0, \sigma^2)$ i.i.d.) where x_i^* can not be measured accurately. Only x_i and y_i with $x_i^* = (1 + \zeta_i)x_i$ are observed, where ζ_i is an unknown measurement error. The least absolute deviations approach estimates β_0 and β_1 by minimizing $\sum_{i=1}^n |y_i - \beta_0 - \beta_1 x_i^*|$:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n \max\{y_i - \beta_0 - \beta_1(1 + \zeta_i)x_i, \beta_0 + \beta_1(1 + \zeta_i)x_i - y_i\}.$$

Because we do not know the values ζ_i, we can apply RO:

$$\min_{\beta_0, \beta_1} \max_{\zeta \in \mathcal{Z}} \sum_{i=1}^n \max\{y_i - \beta_0 - \beta_1(1 + \zeta_i)x_i, \beta_0 + \beta_1(1 + \zeta_i)x_i - y_i\},$$

where we choose $\mathcal{Z}$ to be an ellipsoidal uncertainty region: $\mathcal{Z} = \{\zeta \in \mathbb{R}^n : ||\zeta||_2 \leq \Omega\}$. Note that this is one of the special cases in Section 2.1.3 because it is the sum of absolute values, ζ_i appears only in term i, and the uncertainty region is symmetric around $\zeta = 0$. It follows that the RC can be written as:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n \max\{y_i - \beta_0 - \beta_1 x_i, \beta_0 + \beta_1 x_i - y_i\} + \Omega ||\beta_1 x||_2, \tag{11}$$

which can be reformulated as a CQP. Even though there is an explicit and simple formulation of the exact RC, we also try the other approaches from Section 2 for a comparison. The RC-R is as follows:

$$
\begin{aligned}
\text{(RC-R)} \qquad \min \qquad & \sum_{i=1}^n z_i \\
\text{s.t.} \qquad & z_i \geq y_i - \beta_0 - \beta_1 x_i + \Omega|\beta_1 x_i| \\
& z_i \geq \beta_0 + \beta_1 x_i - y_i + \Omega|\beta_1 x_i|,
\end{aligned}
$$

which for comparison with formulation (11) can also be written as:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n \max\{y_i - \beta_0 - \beta_1 x_i, \beta_0 + \beta_1 x_i - y_i\} + \Omega ||\beta_1 x||_1.$$

This is more pessimistic than constraint (11), since $||\cdot||_1 \geq ||\cdot||_2$. The AARC-R is given by:

$$
\begin{aligned}
\min \quad & d \\
\text{s.t.} \quad & d \geq \sum_{i=1}^n v_i + w_i^\mathsf{T} \zeta && \forall \zeta \in \mathbb{R}^n : ||\zeta||_2 \leq \Omega \\
& v_i + w_i^\mathsf{T} \zeta \geq y_i - \beta_0 - \beta_1(1 + \zeta_i)x_i && \forall \zeta \in \mathbb{R}^n : ||\zeta||_2 \leq \Omega \quad \forall i \in I \\
& v_i + w_i^\mathsf{T} \zeta \geq \beta_0 + \beta_1(1 + \zeta_i)x_i - y_i && \forall \zeta \in \mathbb{R}^n : ||\zeta||_2 \leq \Omega \quad \forall i \in I,
\end{aligned}
$$

which after reformulation becomes:

$$
\begin{aligned}
\text{(AARC-R)} \qquad \min \qquad & d \\
\text{s.t.} \qquad & d \geq \Omega ||w||_2 + \sum_{i=1}^n v_i \\
& v_i \geq y_i - \beta_0 - \beta_1 x_i + \Omega ||\beta_1 x_i e_i + w_i||_2 && \forall i \in I \\
& v_i \geq \beta_0 + \beta_1 x_i - y_i + \Omega ||\beta_1 x_i e_i - w_i||_2 && \forall i \in I,
\end{aligned}
$$

Table 1: A comparison of different solution methods for the regression example averaged over $1,000$ datasets.

Method	v_{method}	$v_{true}(x_{method})$
Nominal	36.326	91.994
RC-R	224.096	92.095
AARC-R	194.091	99.322
QARC-R	99.323	99.322
Exact	91.973	91.973

where e_i is the ith unit vector. The QARC-R is derived in Appendix C.

We have run these four models on $1,000$ cases. In each case, we fixed the parameters at $\beta_0 = 2$, $\beta_1 = 5$ and $\sigma^2 = 1$. A case consists of 15 observations of x_i, generated uniformly in $[0,100]$ i.i.d., $\zeta \in \mathbb{R}^n : ||\zeta||_2 \leq \Omega$ uniformly and $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d. Using these random draws, we have computed $y_i = \beta_0 + \beta_1(1 + \zeta_i)x_i + \varepsilon_i$. For the uncertainty region, we have picked $\Omega = 0.05$.

Next to comparing the usual way by means of v_{true}, we also compare solutions by looking at how well β_0 and β_1 are estimated. Histograms, the mean, and s^2 statistics for β_0, β_1 and v_{true} are shown in Figure 1.

In all $1,000$ cases, the AARC-R and QARC-R give the same estimates of β_0 and β_1. Hence, they are considered equal in both comparisons. All models do well in estimating the parameters β_0 and β_1, except the AARC-R and the QARC-R for β_0.

The objective values (averaged over the $1,000$ cases) are listed in Table 1. The AARC-R and QARC-R objective values are 22.7% respectively 94.4% closer to the optimum than the RC-R objective value. However, for the true value we have $v_{true}(x_{RC-R}) = 92.095$ while $v_{true}(x_{AARC-R}) = v_{true}(x_{QARC-R}) = 99.322$. The RC-R outperforms the AARC-R and the QARC-R in 98% of the cases, and even the nominal solution outperforms the RC-R in 99% of the cases. So, even though the AARC-R and QARC-R provide a much better bound on the optimal true value, their true value is not necessarily better. This shows two things: it may be very misleading to compare robust solutions by their objective values, and using a better approximation of the true objective function might not improve the solution at all.

Algorithms 1 and 2 can run very quickly because the worst case ζ can be computed without running MIP, as this is a simple case. It is therefore more interesting to look at the number of iterations the algorithms takes than at the running time. We have generated 1,000 data sets in which we varied the number of observations. Algorithm 1 adds only 3 or 4 constraints independent of the number of observations. This shows that the method is not only effective for polyhedral uncertainty regions. The number of iterations of Algorithm 2 is shown in Figure 2, where is seems to be a square root function of the number of observations. The regression model $numiter_i = \alpha\sqrt{numobs_i} + \varepsilon_i$ gives $\alpha = 1.258$ with a standard error of 0.007 and $R^2 = 0.966$. For 200 observations, the algorithm generates at most 18 constraints while full EORLC would result in a model with 2^{200} constraints. The algorithm needs more iterations than the vertex enumeration algorithm, but is still very effective.

5.3   Brachytherapy

High dose rate brachytherapy (HDR-BT) is a form of radiation therapy where a highly radioactive sealed source is inserted into a tumour for short time periods via approximately fifteen till twenty catheters. When the catheter positions are fixed, a treatment plan tells for how long the radioactive source has to stay at which position inside the catheters. A perfect treatment plan delivers a prescribed dose to the tumour while not delivering any dose to the surrounding organs at risk. Because this is physically impossible, the goal is to find a good trade-off between irradiating the tumour and saving the organs at risk. The quality of a plan can be measured by means of calculation points. These are artificial points where the received dose can be computed
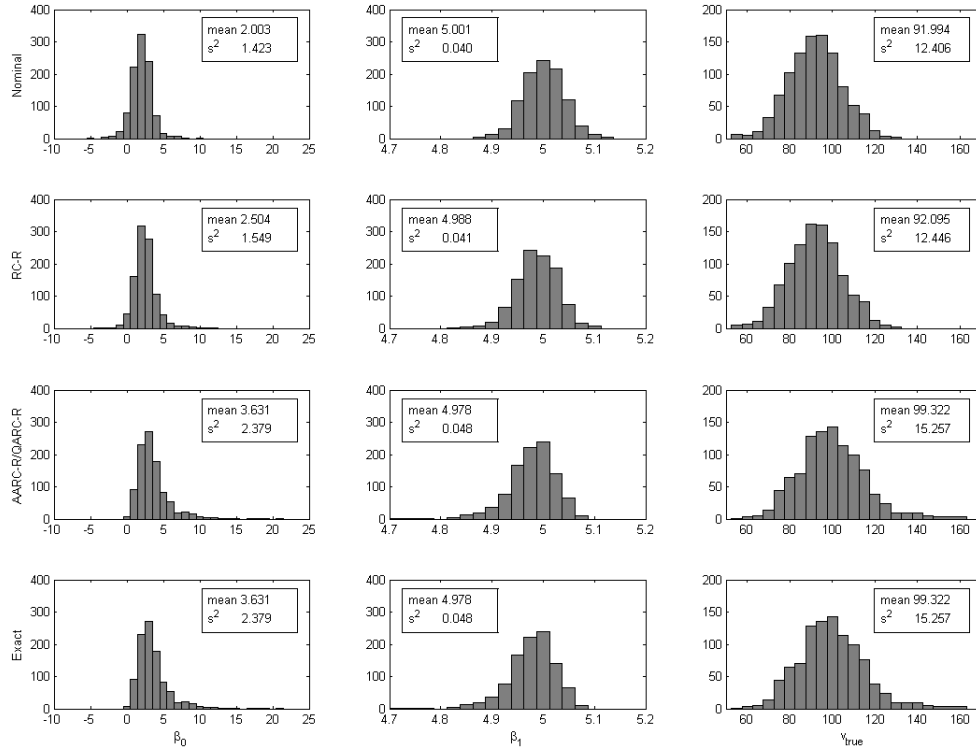
Figure 1: Comparison of the nominal solution, the RC-R solution, the AARC-R/QARC-R solution and the exact solution.
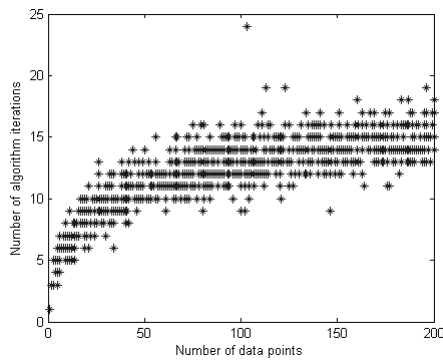

Figure 2: The number of iterations needed by Algorithm 2 to find an optimal solution versus the number of observations in the regression model of Section 5.2.

15

and compared to a prescribed lower and upper bound, that are distributed inside and around the tumour and organs at risk. The dose in a specific calculation point i can be computed as the sum of the individual contributions from every catheter k, which in turn is the sum of the dwell times of the individual dwell positions inside catheter k multiplied by given dose rate vectors d_{ik}:

$$\sum_{k \in K} d_{ik}^{\intercal} t_k.$$

If calculation point i fails to satisfy the lower bound L_i or the upper bound U_i, it contributes a linear penalty of α_i or β_i (respectively) per unit of violation to the objective function. This results in the following optimization problem:

$$\min_{t_k \in \mathbb{R}^n_+} \sum_{i \in I} \max\{0, \alpha_i(L_i - \sum_{k \in K} d_{ik}^{\intercal} t_k), \beta_i(\sum_{k \in K} d_{ik}^{\intercal} t_k - U_i)\}.$$

This convex piecewise linear objective function is commonly used for treatment planning (Alterovitz et al. (2006), Karabis et al. (2009), Lessard and Pouliot (2001)).

The parameters d_{ik} are computed based on the catheter positions. These positions can not be measured accurately, hence the data in the optimization problem is uncertain. We assume that the true catheter position is within some cone around the measured position, which is justified because one side of the catheter is fixed at a known position. We replace the cone with a polyhedral cone with a 10-sided base to end up with a polyhedral set, and we make the simplification that the vector d_{ik} inside the cone is a convex combination of the vectors d_{ik} at the sides of the cone. Thus, we only need to know the vectors d_{ik} at the catheter positions corresponding to the 10 edges of the cone connecting the apex to the base. Using these vectors as the columns of a matrix B, we can write $d_{ik} = B_{ik}\zeta_k$, where $\zeta_k \in \Delta^{|S|-1}$ ($\Delta^{|S|-1} = \{\zeta \in \mathbb{R}^{|S|} : e^{\intercal}\zeta = 1, \zeta \geq 0\}$, the standard simplex in $\mathbb{R}^{|S|}$), and $|S|$ is the number of sides of the base of the polyhedral cone. This gives the following RO problem:

$$
\begin{aligned}
\min \quad & v \\
\text{s.t.} \quad & \sum_i \max\{0, \alpha_i(L_i - \sum_{k \in K}(B_{ik}\zeta_k)^{\intercal} t_k), \\
& \qquad \beta_i(\sum_{k \in K}(B_{ik}\zeta_k)^{\intercal} t_k - U_i)\} \leq v \qquad \forall \zeta_k \in \Delta^{|S|-1}(k \in K) \\
& t_k \geq 0 \qquad\qquad\qquad\qquad\qquad\qquad \forall k \in K.
\end{aligned}
$$

The number of calculation points is usually in the order of magnitude of 5,000. For the purpose of this paper, we have reduced the number of calculation points to 40 because otherwise both the maximization step in Algorithm 2 and the AARC-R are intractable. The model does not lose its value from this reduction, because the objective could be a weighted average between the nominal objective based on all calculation point, and the robust objective based on a small fraction of calculation points that is well distributed inside the tumour.

The RC-R is based on the original constraint with the sum of 40 maxima of 3 functions. We have also solved the RC-R and AARC-R after splitting up the sum, as outlined in the last paragraph of Section 2.2. If we split the sum in groups of size 4, then the number of sums reduces to 10 while each term of sum is the maximum of 3^4 functions.

The results for the different methods are listed in Table 2. The first observation is that the nominal solution has a true value that is five times the optimal value, so the nominal solution is nonrobust. The AARC-R has value ≈ 143, which is much closer to the true value (≈ 137) than to the RC-R value (≈ 229), so the AARC-R almost closes the gap. Algorithm 2 is outperformed by Algorithm 1. Its long running time when MIP is not solved to optimality is not only due to the larger number of iterations, but more importantly, to the gradually increasing running time of solving LP in each iteration. This step becomes so memory consuming, that the algorithm can not finish with CPLEX on 32 bit hardware. Combining both algorithms reduces the number

Table 2: A comparison of different solution methods for the brachytherapy example.

Method	Iterations	Sol. time (s)	v_{method}	$v_{true}(x_{method})$
Nominal	–	0	11.033	711.362
RC-R	–	1	229.986	157.277
AARC-R	–	314	143.104	140.827
RC-R (sum of 8 maxima of 243 functions)	–	51	186.219	157.568
RC-R (sum of 10 maxima of 81 functions)	–	17	203.100	157.030
RC-R (sum of 20 maxima 9 functions)	–	2	218.894	158.438
AARC-R (sum of 10 maxima of 81 functions	–	136214	141.481	140.114
AARC-R (sum of 20 maxima 9 functions)	–	2774	142.447	139.957
Algorithm 1	25	1,173	137.628	137.676
Algorithm 2	36	1,065	137.629	137.662
Combined	20	894	137.628	137.650
Algorithm 1*	53	234	137.627	137.668
Algorithm 2*	> 400	> 50,000	–	–
Combined*	41	353	137.628	137.676

$\varepsilon = 0.05$

* MIP is stopped as soon as the upper bound is at least 0.05 higher than the lower bound

of iterations, as more work per iteration is done, but also increases the solution time in case MIP is not solved to optimality. Splitting up the sum does not give any benefit in terms of true value, though it gives a lower objective value when solving the RC-R.

5.4 Inventory planning

We consider a single item inventory model where backlogging is allowed to compare the AARC-R with the exact RC. At the beginning of each period, an order can be placed that is delivered instantly. At the end of each period, the holding and backlogging costs are c_h and c_b per unit, respectively. The objective is to minimize the costs:

$$\min_{q, w \in \mathbb{R}_+^{|T|}} \max_{d \in \mathcal{Z}} \left(\sum_{t=1}^{|T|} \max\{c_h[x_0 + \sum_{i=1}^t q_i - d_i], c_b[x_0 + \sum_{i=1}^t q_i - d_i]\} \right),$$

where x_0 is the starting inventory, q_i is the order quantity at time i, and d_i is the uncertain demand at time i. In all formulations we allow q_i to depend affinely on the demand in periods 1 up to $i - 1$.

If the uncertainty region is a box, the AARC-R turns out to be exact for the intervals we have tried. This is in accordance with the numbers reported by Ben-Tal et al. (2005). We found that the AARC-R is no longer exact if the uncertainty region is an ellipsoid. Because demand is nonnegative, the ellipsoid is intersected with the nonnegative orthant:

$$\mathcal{Z} = \{d \in \mathbb{R}_+^{|T|} : \left|\left|d - \bar{d}\right|\right|_2 \leq \Omega\}.$$

In our numerical study we have looked at 12 time periods, with parameters $\Omega = 10$, $\bar{d} = 5$, $c_h = 1$, and $c_b = 2$.

Each optimization of MIP in Algorithm 2 takes up to 2 minutes if continued to optimality using CPLEX, making it the most time consuming step in the algorithm. Because the dimension of the uncertain demand vector is 12, a global solver might be faster. We have tried LGO 1.0, whose accuracy can be adjusted with the parameters "maximal number of function evaluations"

Table 3: A comparison of different solution methods for the inventory example.

Method	Iterations	Sol. time (min)	v_{method}	$v_{true}(x_{method})$
Nominal	–	0	0.000	509.903
RC-R	–	0	120.000	94.641
AARC-R	–	0	120.000	93.315
EORLC	–	10	48.750	48.750
RC-R (sum of 2 maxima of 64 functions)	–	0	68.613	54.162
RC-R (sum of 3 maxima of 16 functions)	–	0	83.631	58.140
RC-R (sum of 4 maxima of 8 functions)	–	0	94.456	66.637
RC-R (sum of 6 maxima of 4 functions)	–	0	107.627	77.957
Algorithm 1	123	76	48.749	48.796
Algorithm 2	77	36	48.752	48.802
Combined	79	45	48.755	48.798
Algorithm 1*	309	180	48.752	48.827
Algorithm 2*	112	9	48.753	48.790
Combined *	103	29	48.755	48.825

$\varepsilon = 0.1$

* MIP is stopped as soon as the upper bound is at least 0.1 higher than the lower bound

and "maximal number of stalled evaluations", which initially both are 16,000. Every time the upper bound found by LGO is less than 0.1 larger than the lower bound, the parameters are increased by 25% until they exceed 1,000,000. This reflects the idea that it is still easy to find a violated constraint when the algorithm starts, but gradually becomes more difficult as the quality of the solution increases. Still, CPLEX finds better solutions in less time.

The RC-R is based on the original constraint with the sum of 12 maxima of 2 functions. We have also solved the RC-R and AARC-R after splitting up the sum, as outlined in the last paragraph of Section 2.2. In this example the uncertainty enters the model through time, so the most natural way of splitting up the summation is in consecutive time periods. We split the problem into two groups:

$$\begin{aligned} \min \quad & y_1 + y_2 \\ \text{s.t.} \quad & \sum_{t=1}^{6} \max\{c_h[x_0 + \sum_{i=1}^{t} q_i - d_i], c_b[x_0 + \sum_{i=1}^{t} q_i - d_i] \leq y_1 && \forall d \in \mathcal{Z} \\ & \sum_{t=7}^{12} \max\{c_h[x_0 + \sum_{i=1}^{t} q_i - d_i], c_b[x_0 + \sum_{i=1}^{t} q_i - d_i] \leq y_2 && \forall d \in \mathcal{Z}, \end{aligned}$$

where again we allow q_i to depend affinely on the demand in periods 1 up to $i - 1$. This reformulation introduces more constraints with the sum of maxima, but each sum contains less terms, hopefully resulting in a shorter solution time. The worst case d may differ for the constraints that set y_1 and y_2, hence this reformulation is not exact.

The results are listed in Table 3. Again note that the order policies (q_i) are adjustable in all formulations, including the RC-R, so the differences in this table are caused only by different reformulations of the sum of maxima. The first observation is that the AARC-R gives a very small gain over the RC-R, but still has almost twice the optimal value. So, making the analysis variables adjustable does not significantly improve the solution. Algorithm 2 is the fastest iterative method, requiring approximately the same number of iterations when combined with Algorithm 1. The sum splitting method significantly reduces the computation time at the cost of nonoptimality. It performs much better than the AARC-R, both in optimal value and in true

value. Using the AARC-R on the splitted sums gives a very small gain over the RC-R on the splitted sums, just as for the full problem, so it is not listed in the table. The large true value of the nominal solution comes from the fact that the order sizes are fixed in advance and are not adjusted to observed demand.

6   Conclusions

Because RO is applied constraint-wise, it is very important how constraints are formulated. In this paper we list several approaches to an inequality constraint containing the sum of maxima of linear functions. The RC-R, often used in the literature, is the most pessimistic approach. It is obtained by first reformulating the deterministic constraint into linear constraints using analysis variables, and then applying RO. Its pessimism can be reduced by replacing the analysis variables with linear decision rules before applying RO, which gives the AARC-R. The AARC-R seems to work well for the practical problems we analyzed with polyhedral uncertainty regions, but we have constructed an example with polyhedral uncertainty where the value of the AARC-R is 100% higher than the true value. Nonlinear decision rules may give better results, but are computationally more challenging. The conservatism of the approximations can be reduced by combining several max expressions before reformulating. Especially for ellipsoidal uncertainty this method gives much better solutions at the cost of a slightly higher solution time.

In many cases it is not necessary to use an approximation because an exact reformulation can be practically solved. We identify four special cases in which an exact reformulation is often tractable. For the general case we give two exact general methods: vertex enumeration and EORLC. Both methods may result in very large optimization models, but have an iterative counterpart that can handle larger instances. Vertex enumeration adds a set of constraints for every vertex of the uncertainty region, so this method is preferred if the uncertainty region has a low number of vertices. Surprisingly, its iterative version is also capable of solving problems where the uncertainty region has an infinite number of extreme points efficiently. EORLC is preferred if the number of terms with a maximum function is low. If it is not clear in advance which iterative method is faster, both methods have to be tried because our numerical examples do not show a clear preference. Both methods can be combined, but we have not found a situation in which it is beneficial to do so.

The RC-R is often used in the literature while better approaches could have been applied, mostly without explicitly mentioning that their approach is conservative. In the paper by Kropat and Weber (2008), the exact method EORLC would have increased the number of constraints with only a factor four while reducing the number of variables with almost a factor 2 and not changing the structure of the problem. Bertsimas and Thiele (2006), Wei et al. (2009) and Alem and Morabito (2012) apply the RC-R to a problem with polyhedral uncertainty. Our results, and also the numbers reported by Ben-Tal et al. (2005), suggest that the AARC-R should have been used as it may give good solutions for polyhedral uncertainty regions. For inventory problems in general, when the order quantities are made adjustable, then oftenly also the analysis variables are made adjustable. We show that the latter is not beneficial for ellipsoidal uncertainty, and that both the RC-R and the AARC-R give very bad solutions. For small planning horizons, an exact method has to be used, while for larger horizons splitting the sum in small groups and applying EORLC on the groups significantly improves the solution. Ng et al. (2010) solve a lot allocation problem with ellipsoidal uncertainty. Because the problem is computationally challenging, they solve a problem equivalent to our RC-R using Benders decomposition. Even though all our methods can be applied to their problem, the problem is so challenging that even the simple RC-R can not be solved within a day. We have been able to get a suboptimal solution with iterative vertex enumeration, where both the minimization and the maximization step were stopped before optimality. The solution we got after four hours has a true value of 26.8, whereas the RC-R (which we tried to solve as a MIP) has a value between 28.7 and 47.2. So the objective value of the solution proposed by Ng et al. (2010) is at least 7-76% too pessimistic.

From our numerical examples it becomes clear that the RC-R is not necessarily better than

the nominal problem. Neither of the two optimizes the true problem, so it can not be determined a priori which one has a better true value. The same holds for the AARC-R and the RC-R: If the AARC-R gives a much lower value then at least it provides a guarantee on the worst case, but the RC-R may still outperform the AARC-R. When using an approximation, it is therefore crucial to measure its quality. This can be accomplished by comparing the true value of the solution of the approximation with the value of an exact formulation. If the problem is too large to be solved exactly, the comparison may be based on a smaller instance with similar structure.

A   Derivation of AARC-R using Fenchel's duality

In this appendix we apply Fenchel's duality to robust constraints, a technique introduced in RO by Ben-Tal et al. (2011b). First we will briefly mention the general theory, then we will apply it to constraints of general form, and finally we will apply the general results to constraint (1) and show that the result is the same as the AARC-R.

A.1   Fenchel's duality theorem

We start with some definitions that are necessary to formulate Fenchel's theorem:

Definition A function ϕ is proper convex if it is convex, its codomain is $\mathbb{R} \cup \{\infty\}$, and $\phi(x) < \infty$ for at least one x.

Definition A function ψ is proper concave if it is concave, its codomain is $\mathbb{R} \cup \{-\infty\}$, and $\psi(x) > -\infty$ for at least one x.

Theorem A.1 *(Fenchel's duality (Rockafellar 1970, p. 327)) Let ϕ be a proper convex function on $\mathbb{R}^n$, let ψ be a proper concave function on $\mathbb{R}^n$, and let either of the following conditions be satisifed:*

- *ri(dom ϕ) $\cap$ ri(dom ψ) $\neq \emptyset$*
- *ϕ and ψ are closed, and ri(dom ϕ^*) $\cap$ ri(dom ψ_*) $\neq \emptyset$,*

where ri is the relative interior, dom is the effective domain (dom $\phi = \{x : \phi(x) < \infty\}$), and ϕ^* and ψ_* are the convex and concave conjugate of ϕ and ψ, respectively. That is,

$$\phi^*(s) = \sup_x \{s^\mathsf{T} x - \phi(x)\}$$

$$\psi_*(s) = \inf_x \{s^\mathsf{T} x - \psi(x)\}.$$

Then the following equality holds

$$\inf_x \{\phi(x) - \psi(x)\} = \sup_s \{\psi_*(s) - \phi^*(s)\}.$$

A.1.1   Fenchel's duality applied to a robust constraint of general form.

We focus on the general robust constraint

$$g(\zeta, x) \leq d \qquad \forall \zeta \in \mathcal{Z}, \tag{12}$$

where g is a proper concave function of ζ for any fixed value of x, and the condition for Fenchel's duality is satisfied with respect to the first argument for any fixed value of x. Because values of g are not of interest when $\zeta \notin \mathcal{Z}$, we assume that $g(\zeta, x) = -\infty$ for all $\zeta \notin \mathcal{Z}$. We also assume that $\mathcal{Z}$ is a compact set so that this constraint is equivalent to:

$$\max_{\zeta \in \mathbb{R}^L} \{g(\zeta, x) - \delta_\mathcal{Z}(\zeta)\} \leq d, \tag{13}$$

with $\delta_{\mathcal{Z}}$ the indicator function ($\delta_{\mathcal{Z}}(\zeta) = 0$ if $\zeta \in \mathcal{Z}$, and ∞ otherwise). We can rewrite the left-hand side by applying Fenchel's duality:

$$\min_{s \in \mathbb{R}^L} \{\delta_{\mathcal{Z}}^*(s) - g_*(s, x)\} \leq d,$$

which holds if and only if there exists some $s \in \mathbb{R}^L$ such that:

$$\delta_{\mathcal{Z}}^*(s) - g_*(s, x) \leq d. \tag{14}$$

Note that this constraint is convex in s. Because every step is 'if and only if', constraint (12) is equivalent to constraint (14).

A.1.2 g is the sum of other functions.

If g can be written as the sum of several other functions, it might be impossible or very difficult to find a closed form solution for its conjugate function. Suppose we have a constraint of the form:

$$\sum_{i \in I} g_i(\zeta, x) \leq d \qquad \forall \zeta \in \mathcal{Z}, \tag{15}$$

which is constraint (12) with $g = \sum_{i \in I} g_i$. If we want to formulate an equivalent constraint using Fenchel's duality, we need the concave conjugate of g. Under some mild assumptions on g_i, it turns out to be sufficient to have closed form solutions for the conjugates of g_i. The following lemma appears to be very useful (Rockafellar 1970, p. 145):

Lemma A.2 *Let ψ_i ($i \in I$) be proper concave functions on $\mathbb{R}^n$. If $\cap_{i \in I} ri(dom\ \psi_i) \neq \emptyset$ then*

$$\left(\sum_{i \in I} \psi_i\right)_*(s) = \sup_{\sum_{i \in I} s_i = s} \left\{\sum_{i \in I} (\psi_i)_*(s_i)\right\},$$

and the supremum is attained.

Applying this lemma, we can rewrite constraint (15) as:

$$\delta_{\mathcal{Z}}^*(s) - \max_{\sum_{i \in I} s_i = s} \left\{\sum_{i \in I} (g_i)_*(s_i, x)\right\} \leq d,$$

which is valid if and only if there exists $s_i \in \mathbb{R}^L (i \in I)$ such that

$$\delta_{\mathcal{Z}}^*\left(\sum_{i \in I} s_i\right) - \sum_{i \in I} (g_i)_*(s_i, x) \leq d. \tag{16}$$

So, if all conjugates have closed form solutions, the reformulated constraint also has a closed form. If all conjugates do not have a closed form, this reformulation is still useful because it allows computing each term separately. We will show this for piecewise linear convex functions later in this section. It should again be noted that this constraint is convex in s_i.

A.1.3 f is not convex or g is not concave.

Let us investigate the consequences to the reformulation if g is not concave. We assume g is finite on some non-empty set $\mathcal{Z}$, and ∞ elsewhere, so that its conjugate is non-trivial. Fenchel's inequality (Rockafellar 1970, p. 105) states that:

$$\delta_{\mathcal{Z}}(\zeta) + \delta_{\mathcal{Z}}^*(s) = \delta_{\mathcal{Z}}(\zeta) + \sup_{\zeta' \in \mathcal{Z}} \{\zeta'^\mathsf{T} s - \delta_{\mathcal{Z}}(\zeta')\} \geq \delta_{\mathcal{Z}}(\zeta) + \zeta^\mathsf{T} s - \delta_{\mathcal{Z}}(\zeta) = \zeta^\mathsf{T} s$$

$$g(\zeta, x) + g_*(s, x) = g(\zeta, x) + \inf_{\zeta' \in \mathcal{Z}} \{\zeta'^\mathsf{T} s - g(\zeta', x)\} \leq g(\zeta, x) + \zeta^\mathsf{T} s - g(\zeta, x) = \zeta^\mathsf{T} s,$$

hence:

$$\delta_{\mathcal{Z}}(\zeta) + \delta_{\mathcal{Z}}^*(s) \geq g(\zeta, x) + g_*(s, x),$$

and consequently:

$$g(\zeta, x) - \delta_{\mathcal{Z}}(\zeta) \leq \delta_{\mathcal{Z}}^*(s) - g_*(s, x).$$

This implies that if constraint (14) is satisfied, then so is constraint (13), but the reverse implication is not necessarily true. Hence, constraint (14) is a conservative reformulation of constraint (13).

Also, Lemma A.2 no longer holds with equality. We can rewrite it as an inequality:

$$\left(\sum_{i \in I} \psi_i\right)_*(s) = \inf_t \{s^\mathsf{T} t - \sum_{i \in I} \psi_i(t)\}$$

$$= \inf_t \{\sum_{i \in I} s_i^\mathsf{T} t - \psi_i(t)\},$$

for any $s_1, ..., s_m \in \mathbb{R}^L$ for which $\sum_{i \in I} s_i = s$. So in particular:

$$\left(\sum_{i \in I} \psi_i\right)_*(s) = \sup_{\sum_{i \in I} s_i = s} \{\inf_t \{\sum_{i \in I} s_i^\mathsf{T} t - \psi_i(t)\}\}$$

$$\geq \sup_{\sum_{i \in I} s_i = s} \{\sum_{i \in I} \inf_t \{s_i^\mathsf{T} t - \psi_i(t)\}\}$$

$$= \sup_{\sum_{i \in I} s_i = s} \{\sum_{i \in I} (\psi_i)_*(s_i)\}.$$

This implies that if constraint (16) is satisfied, then so is constraint (15), but the reverse implication is not necessarily true. Hence, constraint (16) can be seen as a conservative reformulation of constraint (15).

A.1.4 g is the sum of pointwise maxima of linear functions.

Let us first derive the conjugates of some functions before we arrive at the theorem:

$$\delta_{\mathcal{Z}}^*(s) = \sup_{\zeta \in \mathbb{R}^L} \{s^\mathsf{T} \zeta - \delta_{\mathcal{Z}}(\zeta)\} = \sup_{\zeta \in \mathcal{Z}} \{s^\mathsf{T} \zeta\} = \max_{\zeta \in \mathcal{Z}} \{s^\mathsf{T} \zeta\},$$

and

$$(\max_{j \in J} \{\ell_{ij}(\zeta, x)\})_*(s_i, x) = \inf_{\zeta \in \mathcal{Z}} \{s_i^\mathsf{T} \zeta - \max_{j \in J} \{\ell_{ij}(\zeta, x)\}\}$$

$$= \inf_{\zeta \in \mathcal{Z}} \{\min_{j \in J} \{s_i^\mathsf{T} \zeta - \ell_{ij}(\zeta, x)\}\}$$

$$= \min_{j \in J} \{\inf_{\zeta \in \mathcal{Z}} \{s_i^\mathsf{T} \zeta - \ell_{ij}(\zeta, x)\}\}.$$

Theorem A.3 *Applying Fenchel's duality to:*

$$\max_{\zeta \in \mathcal{Z}} \sum_{i \in I} \max_{j \in J} \{\ell_{ij}(\zeta, x)\} \leq d, \tag{17}$$

gives a formulation that is equivalent to the AARC-R.

Proof Proof Constraint (17) is equivalent to constraint (15) with $g_i(\zeta, x) = \max_{j \in J} \{\ell_{ij}(\zeta, x)\}$ for ζ in $\mathcal{Z}$ and $g_i(\zeta, x) = -\infty$ otherwise. For for any fixed x, g_i is not concave in ζ so we will end up with a conservative instead of an equivalent reformulation. If we fill in the conjugate functions in constraint (16), the following conservative reformulation is obtained:

$$\max_{\zeta \in \mathcal{Z}} \{\sum_{i \in I} s_i^\mathsf{T} \zeta\} - \sum_{i \in I} \min_{j \in J} \{\inf_{\zeta \in \mathcal{Z}} \{s_i^\mathsf{T} \zeta - \ell_{ij}(\zeta, x)\}\} \leq d.$$

If we model the second terms as $\sum_{i \in I} z_i$, we can write this as:

$$\sum_{i \in I} s_i^\intercal \zeta - \sum_{i \in I} z_i \leq d \qquad \forall \zeta \in \mathcal{Z}$$

$$z_i \leq s_i^\intercal \zeta - \ell_{ij}(\zeta, x) \qquad \forall \zeta \in \mathcal{Z} \quad \forall i \in I \quad \forall j \in J,$$

and by rearranging the terms in each constraint we obtain:

$$\sum_{i \in I} [(-z_i) + s_i^\intercal \zeta] \leq d \qquad \forall \zeta \in \mathcal{Z}$$

$$(-z_i) + s_i^\intercal \zeta \geq \ell_{ij}(\zeta, x) \qquad \forall \zeta \in \mathcal{Z} \quad \forall i \in I \quad \forall j \in J,$$

which is the same as the AARC-R. ∎

B   Derivation of AARC-R by reformulating the nonrobust constraint

In this appendix we give a different derivation of the AARC-R of constraint (1) when both the biaffine functions and the uncertainty region are separable in the following way:

$$\sum_{i \in I} \max_{j \in J} \{ \sum_{k \in K} \ell_{ijk}(\zeta_k, x) \} \leq d \qquad \forall \zeta_k \in \mathcal{Z}_k \quad (k \in K), \tag{18}$$

and $\mathcal{Z}_k$ is the convex hull of different scenarios ζ_k^s ($s \in S$). An example where this constraint is commonly used, is HDR brachytherapy optimization (Alterovitz et al. (2006), Karabis et al. (2009), Lessard and Pouliot (2001)). If the summation over k were outside the *max* expression, then an analysis variable could be used for every k without introducing any conservatism. Vertex enumeration can then be done on every $\mathcal{Z}_k$ separately. For problems not affected by uncertainty, we generalize a formulation trick by Balvert (2011) to show that indeed there is an equivalent formulation where the summation over k is outside the *max* expression. Then we show equivalence to the AARC-R if there is uncertainty. First, we prove the following equality for fixed x and ζ_k:

Lemma B.1

$$\sum_{i \in I} \max_{j \in J} \{ \sum_{k \in K} \ell_{ijk}(\zeta_k, x) \} = \min_{y \in \mathbb{R}^{|I||J||K|}: \sum_{k \in K} y_{ijk} = 0} \sum_{k \in K} \sum_{i \in I} \max_{j \in J} \{ y_{ijk} + \ell_{ijk}(\zeta_k, x) \}. \tag{19}$$

Proof   Proof Note that:

$$\sum_{k \in K} \sum_{i \in I} \max_{j \in J} \{ y_{ijk} + \ell_{ijk}(\zeta_k, x) \} \geq \sum_{k \in K} \sum_{i \in I} y_{ij(i)k} + \ell_{ij(i)k}(\zeta_k, x) = \sum_{i \in I} \sum_{k \in K} \ell_{ij(i)k}(\zeta_k, x) \quad \forall j(i) \in J,$$

for any $j(i)$ in J, so in particular:

$$\sum_{k \in K} \sum_{i \in I} \max_{j \in J} \{ y_{ijk} + \ell_{ijk}(\zeta_k, x) \} \geq \sum_{i \in I} \max_{j \in J} \{ \sum_{k \in K} \ell_{ijk}(\zeta_k, x) \},$$

for any y, so in particular for the minimum. Hence, the right hand side of (19) is as least as large as the left hand side. On the other hand, given a feasible point for the left hand side of (19), we can always construct a feasible point for the right hand side with equal value by taking the same x, and $y_{ijk} = \frac{1}{|K|} \sum_{k' \in K} \ell_{ijk'}(\zeta_{k'}, x) - \ell_{ijk}(\zeta_k, x)$:

$$\sum_{k \in K} \sum_{i \in I} \max_{j \in J} \{ y_{ijk} + \ell_{ijk}(\zeta_k, x) \} = \sum_{i \in I} \sum_{k \in K} \max_{j \in J} \{ \frac{1}{|K|} \sum_{k' \in K} \ell_{ijk'}(\zeta_{k'}, x) \} = \sum_{i \in I} \max_{j \in J} \{ \sum_{k' \in K} \ell_{ijk'}(\zeta_{k'}, x) \}.$$

∎

Lemma B.2 *A conservative reformulation of constraint* (18) *is given by:*

$$\sum_{k \in K} \sum_{i \in I} \max_{j \in J} \{y_{ijk} + \ell_{ijk}(\zeta_k, x)\} \le d \qquad \forall \zeta_k \in \mathcal{Z}_k \quad (k \in K)$$

$$\sum_{k \in K} y_{ijk} = 0 \qquad \forall i \in I \quad \forall j \in J.$$

Proof Proof We use the proof of Lemma B.1. The first part of the proof still holds, so the equality in constraint (19) becomes a "$\le$". The construction of the feasible point in the second part of the proof depends on the value $\ell_{ijk}(\zeta_k, x)$. This value does not exist in the robust constraint (18), since there is no single ζ_k. Note that the left hand sides of (18) and (19) are the same, so by replacing the left hand side of (18) with the right hand side of (19), the conservative reformulation is obtained. ∎

The uncertainty is separable per k in the first constraint of the conservative reformulation. Hence, analysis variables per k can replace each term without changing the solution. Vertex enumeration can then be done for every $\mathcal{Z}_k$ separately:

$$\sum_{k \in K} z_k \le d \qquad (20)$$

$$z_k \ge \sum_{i \in I} w_{iks} \qquad \forall s \in S$$

$$w_{iks} \ge y_{ijk} + \ell_{ijk}(\zeta_k^s, x)\} \qquad \forall i \in I \quad \forall s \in S \quad \forall k \in K$$

$$\sum_{k \in K} y_{ijk} = 0 \qquad \forall i \in I \quad \forall j \in J.$$

It remains to show that this formulation is equivalent to the AARC-R of constraint (18).

Theorem B.3 *The conservative reformulation in Lemma B.2 is the AARC-R in case of scenario generated uncertainty.*

Proof Proof The AARC-R of constraint (18) is given by:

$$\sum_{i \in I} \left(v_i + \sum_{k \in K} w_{ik}^{\mathsf{T}} \zeta_k \right) \le d \qquad \forall \zeta_k \in \mathcal{Z}_k \quad (k \in K)$$

$$v_i + \sum_{k \in K} w_{ik}^{\mathsf{T}} \zeta_k \ge \sum_{k \in K} \ell_{ij}(\zeta_k, x) \qquad \forall i \in I \quad \forall j \in J \quad \forall \zeta_k \in \mathcal{Z}_k \quad (k \in K). \qquad (21)$$

Constraint (21) can be reformulated as:

$$w_{ik}^{\mathsf{T}} \zeta_k \ge y_{ijk} + \ell_{ij}(\zeta_k, x) \qquad \forall i \in I \quad \forall j \in J \quad \forall \zeta_k \in \mathcal{Z}_k \quad \forall k \in K$$

$$\sum_{k \in K} y_{ijk} = -v_i \qquad \forall i \in I \quad \forall j \in J.$$

We have assumed scenario generated uncertainty, so our uncertainty region is $\mathcal{Z}_k = \Delta^{|S|-1}$, the standard simplex in $\mathbb{R}^{|S|}$. Suppose an optimal solution has $v_i \ne 0$, then an optimal solution with $v_i = 0$ can be obtained by increasing all elements of w_{ik} for a single random k with v_i because the elements of ζ_k sum to 1. Hence we fix $v_i = 0$. The AARC-R can then be formulated as:

$$\sum_{k \in K} z_k \le d$$

$$z_k \ge \sum_{i \in I} w_{ik}^{\mathsf{T}} \zeta_k \qquad \forall \zeta_k \in \mathcal{Z}_k \quad (k \in K)$$

$$w_{ik}^{\mathsf{T}} \zeta_k \ge y_{ijk} + \ell_{ij}(\zeta_k, x) \qquad \forall i \in I \quad \forall j \in J \quad \forall \zeta_k \in \mathcal{Z}_k \quad \forall k \in K$$

$$\sum_{k \in K} y_{ijk} = 0 \qquad \forall i \in I \quad \forall j \in J.$$

Let w_{iks} denote the sth component of w_{ik}. Equivalence to (20) now follows from vertex enumeration. ∎

C Derivation of the QARC-R for an ellipsoidal uncertainty region

In this appendix the SDP reformulation of the QARC-R of (1) is derived for an ellipsoidal uncertainty region. Constraint (2) in the QARC-R is a quadratic constraint in ζ, so there exists $v : \mathbb{R}^n \to \mathbb{R}$, $w : \mathbb{R}^n \to \mathbb{R}^L$ and $W : \mathbb{R}^n \to \mathbb{S}^L$ (the space of symmetric matrices of size $L \times L$) such that the constraint can be written as:

$$v(x) + w(x)^\mathsf{T}\zeta + \zeta^\mathsf{T}W(x)\zeta \leq d \qquad\qquad \forall \zeta \in \mathbb{R}^L : ||\zeta||_2 \leq \Omega. \qquad (22)$$

This constraint is valid if and only if:

$$v(x) + \max_{\zeta \in \mathbb{R}^L : ||\zeta||_2 \leq \Omega} \{w(x)^\mathsf{T}\zeta + \zeta^\mathsf{T}W(x)\zeta\} \leq d. \qquad (23)$$

The maximization problem at the left hand side is equivalent to an SDP (Ben-Tal et al. 2009a, p. 382). In order to simplify notation, let:

$$C = \begin{pmatrix} 0 & \frac{1}{2}v(x)^\mathsf{T} & 0 \\ \frac{1}{2}v(x) & W(x) & 0 \\ 0 & 0^\mathsf{T} & 0 \end{pmatrix}$$

After introducing a slack variable X_{33}, the maximization problem can be written as:

$$\begin{aligned}
\max \quad & \langle C, X \rangle \\
\text{s.t.} \quad & \operatorname{tr} X = 1 + \Omega^2 \\
& X = \begin{pmatrix} 1 & X_{12}^\mathsf{T} & 0 \\ X_{12} & X_{22} & 0 \\ 0 & 0 & X_{33} \end{pmatrix} \succeq 0 \\
& X_{12} \in \mathbb{R}^L, X_{22} \in \mathbb{R}^{L \times L}, X_{33} \in \mathbb{R},
\end{aligned}$$

where X_{12} can be seen as a replacement for ζ, X_{22} as a replacement for $\zeta\zeta^\mathsf{T}$, and $\langle \cdot, \cdot \rangle$ denotes the trace inner product. Let z_1^0 be the dual variable associated with the constraint $\operatorname{tr} X = 1+\Omega^2$, and let z_i^0 ($i = 2,3,4$) be the dual variables corresponding to the constraints $X_{11} = 1$, $X_{1,n+2} = 0$, etc. The dual problem is given by:

$$\begin{aligned}
\min \quad & (1 + \Omega^2)z_1^0 + z_2^0 \\
\text{s.t.} \quad & \begin{pmatrix} z_1^0 + z_2^0 & 0^\mathsf{T} & z_3^0 \\ 0 & z_1^0 I & z_4^0 \\ z_3^0 & z_4^{0\,\mathsf{T}} & z_1^0 \end{pmatrix} - C \succeq 0 \\
& z_1^0 \in \mathbb{R}, z_2^0 \in \mathbb{R}, z_3^0 \in \mathbb{R}, z_4^0 \in \mathbb{R}^L,
\end{aligned}$$

where I is the $L \times L$ identity matrix. After substituting this into constraint (23) and removing the min operator, we get the following reformulation of constraint (22):

$$\begin{aligned}
& v(x) + (1 + \Omega^2)z_1^0 + z_2^0 \leq d \\
& \begin{pmatrix} z_1^0 + z_2^0 & 0^\mathsf{T} & z_3^0 \\ 0 & z_1^0 I & z_4^0 \\ z_3^0 & z_4^{0\,\mathsf{T}} & z_1^0 \end{pmatrix} - C \succeq 0 \\
& z_1^0 \in \mathbb{R}, z_2^0 \in \mathbb{R}, z_3^0 \in \mathbb{R}, z_4^0 \in \mathbb{R}^L.
\end{aligned}$$

Similarly, constraint (3) in the QARC-R is a quadratic constraint in ζ, so there exists $v^{ij} : \mathbb{R}^n \to \mathbb{R}$, $w^{ij} : \mathbb{R}^n \to \mathbb{R}^L$ and $W^{ij} : \mathbb{R}^n \to \mathbb{S}^L$ (the space of symmetric matrices of size $L \times L$) such that the constraint can be written as:

$$v^{ij}(x) + w^{ij}(x)^\intercal \zeta + \zeta^\intercal W^{ij}(x)\zeta \geq 0 \qquad\qquad \forall \zeta \in \mathbb{R}^L : ||\zeta||_2 \leq \Omega,$$

which can be rewritten as:

$$v^{ij}(x) + (1 + \Omega^2)z_1^{ij} + z_2^{ij} \geq 0$$
$$-\begin{pmatrix} z_1^{ij} + z_2^{ij} & -\frac{1}{2}v^{ij}(x)^\intercal & z_3^{ij} \\ -\frac{1}{2}v^{ij}(x) & z_1^{ij}I - W^{ij}(x) & z_4^{ij} \\ z_3^{ij} & z_4^{ij\,\intercal} & z_1^{ij} \end{pmatrix} \succeq 0$$
$$z_1^{ij} \in \mathbb{R}, z_2^{ij} \in \mathbb{R}, z_3^{ij} \in \mathbb{R}, z_4^{ij} \in \mathbb{R}^L.$$

The QARC-R has $|I||J| + 1$ matrices of size $|L + 2| \times |L + 2|$ and $|I||J| + 1$ constraints.

References

Alem, D. J., R. Morabito. 2012. Production planning in furniture settings via robust optimization. *Computers & Operations Research* **39** 139 – 150. URL http://dx.doi.org/10.1016/j.cor.2011.02.022.

Alterovitz, R., E. Lessard, J. Pouliot, I. J. Hsu, J. F. O'Brien, K. Goldberg. 2006. Optimization of HDR brachytherapy dose distributions using linear programming with penalty costs. *Medical Physics* **33** 4012–4019. URL http://dx.doi.org/10.1118/1.2349685.

Balvert, M. 2011. Robustness of treatment planning against uncertainty in catheter positions in HDR brachytherapy. BSc thesis, Tilburg University.

Ben-Tal, A., B. D. Chung, S. R. Mandala, T. Yao. 2011a. Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation Research Part B: Methodological* **45** 1177 – 1189. URL http://dx.doi.org/10.1016/j.trb.2010.09.002.

Ben-Tal, A., D. den Hertog, J-Ph. Vial. 2011b. Deriving robust counterparts of nonlinear uncertain inequalities. Working paper.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009a. *Robust Optimization*. Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ. URL http://sites.google.com/site/robustoptimization/.

Ben-Tal, A., B. Golany, A. Nemirovski, J-Ph. Vial. 2005. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management* **7** 248–271. URL http://dx.doi.org/10.1287/msom.1050.0081.

Ben-Tal, A., B. Golany, S. Shtern. 2009b. Robust multi-echelon multi-period inventory control. *European Journal of Operational Research* **199** 922–935. URL http://dx.doi.org/10.1016/j.ejor.2009.01.058.

Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99** 351–376. URL http://dx.doi.org/10.1007/s10107-003-0454-y.

Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations Research Letters* **25** 1–13. URL http://dx.doi.org/10.1016/S0167-6377(99)00016-4.

Bertsimas, D., D. A. Iancu, P. A. Parrilo. 2009. A hierarchy of near-optimal policies for multi-stage adaptive optimization. *To appear in IEEE Automatic Control* URL http://www.optimization-online.org/DB_FILE/2009/06/2327.pdf.

Bertsimas, D., A. Thiele. 2004. A robust optimization approach to supply chain management. Daniel Bienstock, George Nemhauser, eds., *Integer Programming and Combinatorial Optimization*, *Lecture Notes in Computer Science*, vol. 3064. Springer Berlin / Heidelberg, 145–156. URL http://dx.doi.org/10.1007/978-3-540-25960-2_7.

Bertsimas, D., A. Thiele. 2006. A robust optimization approach to inventory theory. *Operations Research* **54** 150–168. URL http://dx.doi.org/10.1287/opre.1050.0238.

Bienstock, D., N. Özbay. 2008. Computing robust basestock levels. *Discrete Optimization* **5** 389–414. URL http://dx.doi.org/10.1016/j.disopt.2006.12.002.

Bohle, C., S. Maturana, J. Vera. 2010. A robust optimization approach to wine grape harvesting scheduling. *European Journal of Operational Research* **200** 245–252. URL http://dx.doi.org/10.1016/j.ejor.2008.12.003.

El Ghaoui, L., H. Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications* **18** 1035–1064. URL http://dx.doi.org/10.1137/S0895479896298130.

Grant, M., S. Boyd. 2010. CVX: Matlab software for disciplined convex programming, version 1.21. http://cvxr.com/cvx.

Karabis, A., P. Belotti, D. Baltas. 2009. Optimization of catheter position and dwell time in prostate hdr brachytherapy using hipo and linear programming. *World Congress on Medical Physics and Biomedical Engineering*, *Germany IFMBE Proceedings*, vol. 25/I. 612–615.

Kropat, E., G.-W. Weber. 2008. Robust regression analysis for gene-environment and eco-finance networks. URL http://www3.iam.metu.edu.tr/iam/images/3/32/Preprint184.pdf. Working Paper.

Lessard, E., J. Pouliot. 2001. Inverse planning anatomy-based dose optimization for HDR-brachytherapy of the prostate using fast simulated annealing algorithm and dedicated objective function. *Medical Physics* **28** 773–779.

Löfberg, J. 2010. Automatic robust convex programming. *Optimization methods and software* URL http://dx.doi.org/10.1080/10556788.2010.517532. Accepted for publication.

Ng, T. S., Y. Sun, J. Fowler. 2010. Semiconductor lot allocation using robust optimization. *European Journal of Operational Research* **205** 557–570. URL http://dx.doi.org/10.1016/j.ejor.2010.01.021.

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

Soyster, A. L. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* **21** 1154–1157.

Thiele, A. 2004. *A robust optimization approach to supply chains and revenue management*. URL http://dspace.mit.edu/handle/1721.1/16693. PhD thesis, Massachusetts Institute of Technology.

Wei, C., Y. Li, X. Cai. 2009. Robust optimal policies of production and inventory with uncertain returns and demand. *International Journal of Production Economics* –URL http://dx.doi.org/10.1016/j.ijpe.2009.11.008.

Xu, H., C. Caramanis, S. Mannor. 2009. Robustness and regularization of support vector machines. *The Journal of Machine Learning Research* **10** 1485–1510.

Yanıkoğlu, İ., A. Ben-Tal, D. den Hertog, M. Laurent. 2011. New nonlinear decision rules for adjustable robust counterparts Work in progress.