Approximation of hard uncertain convex inequalities

Ernst Roos[*], Dick den Hertog
CentER and Department of Econometrics and Operations Research, Tilburg University

Aharon Ben-Tal
Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology
Shenkar College, Israel

Frans J.C.T. de Ruiter
CQM, Eindhoven, The Netherlands

Jianzhe Zhen
Risk Analytics and Optimization Chair,
École Polytechnique Fédérale de Lausanne, Switzerland

February 27, 2019

Abstract

Robust Optimization is a widespread approach to treat uncertainty in optimization problems. Finding a computationally tractable formulation of the robust counterpart of an uncertain optimization problem is a key step in applying this approach. Techniques for finding a computationally tractable robust counterpart are available for constraints concave in the uncertain parameters. In many problems, however, the uncertain parameters appear in a convex way, which is problematic as no general techniques exist for such problems. In this paper, we provide a systematic way to construct safe approximations to such constraints when the uncertainty set is polyhedral. Specifically, we reformulate the original problem as a linear adjustable robust optimization problem in which the nonlinearity of the original problem is captured by the new uncertainty set. Additionally, we prove that preprocessing a constraint with a concave transformation that preserves its convexity can tighten the safe approximation obtained. We subsequently apply our theory to quadratic constraints, constraints that are the sum of maxima and the sum of maxima squared, as well as constraints from geometric programming. We demonstrate the quality of the approximations with a study of geometric programming problems and numerical examples from radiotherapy optimization, which contain a constraint of the sum of maxima squared type.

[*]Corresponding author: e.j.roos@tilburguniversity.edu

1 Introduction

In practice, optimization problems frequently contain some form of uncertainty. This uncertainty arises as a result of prediction, estimation or implementation errors, for example. Robust Optimization (RO) is one of the main approaches to address such uncertainty and was initiated by Ben-Tal and Nemirovski (1998) and El Ghaoui and Lebret (1997).

Whereas other approaches, such as Stochastic Programming (Shapiro et al., 2014) or Distributionally Robust Optimization (see e.g. Wiesemann et al. (2014)), assume that the uncertain parameters are random variables with fully or partially known distribution functions, RO models uncertainty by means of an uncertainty set. This set contains all possible scenarios for the uncertain parameters one wants to safeguard against.

One of the main advantages of robust optimization is the computational tractability of the resulting optimization problems. Indeed, for many types of robust problems it has been shown that its robust counterpart is tractable. Ben-Tal et al. (2009) give a comprehensive account of robust optimization techniques to obtain such tractable formulations. A general approach to finding computationally tractable reformulations is presented in Ben-Tal et al. (2015), where a technique is presented that is applicable to any constraint that is convex in the optimization variables and concave in the uncertain parameters. This technique does not necessarily lead to a closed-form robust counterpart for any convex uncertainty set, but does present solutions for the commonly considered uncertainty sets like polyhedral and ellipsoidal uncertainty.

In contrast, obtaining a tractable robust counterpart for constraints that are *convex* in the uncertain parameters is, in general, hard (Chassein and Goerigk, 2019). Such constraints are, however, common; they appear, for example, in inventory management problems, geometric programming and conic optimization. For specific types of constraints such as second-order cone (SOC) and semidefinite programming (SDP) constraints combined with specific uncertainty sets, e.g., norm-bounded or polyhedral uncertainty sets, computationally tractable approximations exist. Some of the first techniques addressing these problems are discussed in El Ghaoui and Lebret (1997); El Ghaoui et al. (1998); Ben-Tal et al. (2002); Bertsimas and Sim (2006) and a more recent technique is discussed in Zhen et al. (2017).

In this paper, we consider general constraints that are convex in both the optimization variables and the uncertain parameters where the latter reside in polyhedral uncertainty sets. More specifically, we convert the robust counterpart to an equivalent linear adjustable robust optimization problem. This enables us to treat the converted problem with existing techniques for Adjustable Robust Optimization (ARO), such as (non)linear decision rules and Fourier-Motzkin elimination. In particular, after substituting linear decision rules, the resulting problem is a safe approximation to the true robust counterpart.

Our results are a generalization of the results derived by Zhen et al. (2017). They use a similar idea of reformulation to an adjustable robust optimization problem such that safe approximations can be constructed based on existing ARO techniques. The approach they describe, however, is only applicable to second-order cone and semidefinite programming constraints, while the approach we propose is much more general, as it can be applied to a broad class of constraints that are convex in both the uncertain parameters and the decision variables.

Additionally, we treat possible alternative formulations to the original problem and discuss the potential differences in the safe approximation our approach yields for said approximations. We show that applying a concave transformation that preserves the convexity of the original constraint yields an approximation that is at least as tight as applying our approach to the original constraint.

We discuss the application of our approach to four important classes of constraints for which no computationally tractable exact reformulation of the robust counterpart is known: quadratic, sum-of-max, sum-of-max squared and log-sum-exp constraints. In particular, we discuss how our approach and the one proposed by Zhen et al. (2017) compare for quadratic constraints, which can be equivalently stated as second-order cone constraints. We also provide numerical results for radiotherapy optimization problems containing a sum-of-max squared constraint and geometric programming problems that contain log-sum-exp constraints. These numerical experiments illustrate the quality of the safe approximation provided in this paper.

The paper is organized as follows. Section 2 treats the general reformulation we propose, as well as possible alternative formulations and methods to solve said reformulation. Section 3 treats four theoretical examples: quadratic constraints, sum-of-max constraints, sum-of-max squared constraints and log-sum-exp constraints. Section 4 contains numerical results for geometric programming and radiotherapy optimization.

Notation. Throughout this paper we use the following notation.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a closed convex function with domain $\mathrm{dom}(f) = \{x \mid f(x) < \infty\}$. The convex *conjugate*, which we refer to as conjugate, of f is defined as

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \left\{ y^\top x - f(x) \right\}.$$

The *perspective function* of f is defined by

$$(f\lambda)(x) = \begin{cases} \lambda f\left(\frac{x}{\lambda}\right) & \lambda > 0 \\ f_\infty(x) & \lambda = 0, \end{cases}$$

which implies $(f\lambda)$ is a closed convex function, and where f_∞ is the *recession function* of

f, defined by (Rockafellar, 1970, Corollary 8.5.2):

$$f_\infty(y) = \lim_{\lambda \downarrow 0} \lambda f\left(\frac{y}{\lambda}\right). \tag{1}$$

We write $\lambda f\left(\frac{x}{\lambda}\right)$ for the perspective function throughout the rest of the paper, implicitly assuming that for $\lambda = 0$, we use $f_\infty(x)$.

The *support function* of a set U is the conjugate of that set's *indicator function*. This indicator function defined as:

$$\delta\left(x \mid U\right) = \begin{cases} 0 & \text{if } x \in U \\ \infty & \text{otherwise,} \end{cases}$$

and thus the support function is given by

$$\delta^*\left(y \mid U\right) = \sup_{x \in U} y^\top x.$$

This support function paves the way for an alternative definition of the recession function of f, which we also use (Rockafellar, 1970, Theorem 13.3):

$$f_\infty(y) = \delta^*\left(y \mid \mathrm{dom}\left(f^*\right)\right). \tag{2}$$

For $f(x) = \sum_{i=1}^n \sqrt{1 + x_i^2}$, for example, the recession function is equal to $f_\infty(y) = \|y\|_2$. An in-depth analysis of the recession function, including a table of recession functions of some well-known functions, can be found in Appendix A.

2 The Robust Counterpart

2.1 Reformulation to ARO

In this paper, we consider a general convex constraint given by

$$f\left(A(x)\zeta + b(x)\right) \leq 0 \qquad \forall \zeta \in U, \tag{P}$$

where $f : \mathbb{R}^p \to \mathbb{R}$ is convex and closed, $A : \mathbb{R}^n \to \mathbb{R}^{p \times L}$ and $b : \mathbb{R}^n \to \mathbb{R}^p$ are affine, and U is a nonempty polyhedron given by

$$U = \left\{\zeta \in \mathbb{R}_+^L \;\middle|\; D\zeta = d\right\}, \tag{3}$$

for some $D \in \mathbb{R}^{q \times L}$ and $d \in \mathbb{R}^q$.

This formulation allows for many important classes of constraints, such as quadratic, sum-of-max (with and without square) and log-sum-exp functions, which we discuss in Section 3. Other examples are (sums of) norms and negative entropy. Furthermore, all functions $g(x,\zeta)$ that are jointly convex in x and ζ can be written as $f\left(A(x)\zeta + b(x)\right)$ for some affine functions $A(x)$ and $b(x)$.

A direct implication of this remark is that we can also find safe approximations for constraints

$$g\left(x, y(\zeta)\right) \leq 0 \qquad \forall \zeta \in U, \tag{4}$$

where g is jointly convex in x and y for adjustable variables y, that is, jointly convex constraints in adjustable robust optimization problems. Specifically, such constraints can be treated by substituting a linear decision rule $y = s + S\zeta$, such that $g\left(x, s + S\zeta\right)$ can be written as $f\left(A(s,S,x)\zeta + b(s,S,x)\right)$. We do remark that substituting a linear decision rule for y yields a safe approximation to (4), and thus our approach yields a safe approximation to this safe approximation of (4).

We also note that by choosing to consider the function $f(A(x)\zeta + b(x))$ we cannot handle all constraints convex in both x and ζ. Functions of the form $b(x)^\top g(\zeta)$, for an affine function b and convex function g, cannot, for example, be formulated as $f(A(x)\zeta + b(x))$. An example of such functions occurs in brachytherapy optimization (Gorissen et al., 2013). Other examples include, but are not limited to capital budgeting problems and multinomial logit models (Alfandari and García, 2018).

It is also important to discuss the added value of our approach for a more general convex function over the previously developed approaches for second-order cone programming (SOCP) and semidefinite programming (SDP) constraints by Zhen et al. (2017). Although many (nominal) problems can be reformulated as an SOCP, the robust counterparts of these equivalent problems are not necessarily the same. As an example, we consider a constraint on the sum of maxima squared:

$$\sum_{i=1}^{p} \max\left\{A_i(x)\zeta + b_i(x),\ 0\right\}^2 \leq 0 \qquad\qquad \forall \zeta \in U, \tag{5}$$

and its SOCP reformulation

$$\forall \zeta \in U : \begin{cases} \|y(\zeta)\|_2 \leq 0 & & \text{(6a)} \\ A_i(x)\zeta + b_i(x) \leq y_i(\zeta) & i = 1,\ldots,p & \text{(6b)} \\ 0 \leq y_i(\zeta) & i = 1,\ldots,p. & \text{(6c)} \end{cases}$$

While the non-robust versions of (5) and (6) are clearly equivalent, (5) and (6) are in fact only equivalent if the newly introduced variables y_i are considered to be adjustable (Gorissen et al., 2015). As existing techniques for SOCP problems with uncertainty do not allow for adjustable variables, some form of (linear) decision rule should be substituted for $y_i(\zeta)$ in order to apply them. Such an approach would thus lead to constructing a safe approximation of the safe approximation of (6) obtained by substituting the decision rules. Our approach, on the other hand, can be applied to (5) directly, eliminating one of these layers of approximation.

In order to find a safe approximation to (P), we first transform the problem into an equivalent linear adjustable robust optimization problem.

Theorem 1. *Let $f : \mathbb{R}^p \to \mathbb{R}$ be a closed convex function and let $A : \mathbb{R}^n \to \mathbb{R}^{p\times L}$ and $b : \mathbb{R}^n \to \mathbb{R}^p$ be affine functions. Let $U \subseteq \mathbb{R}^L$ be a polyhedron as defined in* (3)*. Then, $x \in \mathbb{R}^n$ satisfies*

$$f\left(A(x)\zeta + b(x)\right) \leq 0 \qquad \forall \zeta \in U,$$

if and only if it satisfies the following set of adjustable robust optimization constraints:

$$\forall w \in \operatorname{dom}\left(f^*\right),\ \exists \lambda \in \mathbb{R}^q : \begin{cases} d^\top\lambda + w^\top b(x) - f^*(w) \leq 0 & \text{(7a)} \\ D^\top\lambda \geq A(x)^\top w. & \text{(7b)} \end{cases}$$

Proof. Because f is a closed convex function we have that

$$f(z) = f^{**}(z) = \sup_{w\in\operatorname{dom} f^*} \left\{ z^\top w - f^*(w) \right\}.$$

Substituting this into (P) yields

$$\begin{aligned}
& \forall \zeta \in U \ :\ f\left(A(x)\zeta + b(x)\right) \leq 0 \\
\iff\quad & \forall \zeta \in U \ :\ \sup_{w\in\operatorname{dom} f^*} \left\{ \left(A(x)\zeta + b(x)\right)^\top w - f^*(w) \right\} \leq 0 \\
\iff\quad & \sup_{\zeta\in U} \left\{ \sup_{w\in\operatorname{dom} f^*} \left\{ \left(A(x)^\top w\right)^\top \zeta + b(x)^\top w - f^*(w) \right\} \right\} \leq 0 \\
\iff\quad & \sup_{w\in\operatorname{dom} f^*} \left\{ \sup_{\zeta\in U} \left\{ \left(A(x)^\top w\right)^\top \zeta \right\} + b(x)^\top w - f^*(w) \right\} \leq 0 \qquad (8) \\
\iff\quad & \sup_{w\in\operatorname{dom} f^*} \left\{ \inf_{\lambda\in\mathbb{R}^q} \left\{ d^\top\lambda \mid D^\top\lambda \geq A(x)^\top w \right\} + b(x)^\top w - f^*(w) \right\} \leq 0 \qquad (9) \\
\iff\quad & \forall w \in \operatorname{dom} f^*,\ \exists \lambda \in \mathbb{R}^q \ :\ \begin{cases} d^\top\lambda + b(x)^\top w - f^*(w) \leq 0 \\ D^\top\lambda \geq A(x)^\top w \end{cases}
\end{aligned}$$

where (8) and (9) are equivalent because of strong LP duality. $\square$

The equivalent formulation given in (7) can be interpreted as a set of linear adjustable robust optimization constraints, because it states that for each value of w there must exist a value λ that satisfies the constraints, that is, the value of λ can depend on w. In the field of Robust Optimization such a variable λ is referred to as an adjustable variable, as its value can be adjusted after the value of the uncertain parameter w is revealed.

We note that a similar result holds if the nonnegativity constraint in U is omitted. Then the inequality in (7b) becomes an equality constraint, which can be used to eliminate some adjustable variables λ. Eliminating variables in this way is equivalent to imposing linear decision rules (Zhen and Den Hertog, 2017, Lemma 2).

2.2 Safe Approximation

Since the equivalent formulation derived in Theorem 1 is a set of adjustable robust linear constraints, conventional techniques for such problems can be applied. A common technique to obtain a computationally tractable safe approximation to adjustable robust constraints is limiting the adjustable variables to be linear decision rules in the uncertain parameters (Ben-Tal et al., 2004).

Theorem 2. *If there exist $u \in \mathbb{R}^q$ and $V \in \mathbb{R}^{q\times p}$ for a given $x \in \mathbb{R}^n$ such that*

$$\begin{cases} d^\top u + f\left(b(x) + V^\top d\right) \le 0 & \qquad (10a) \\ f_\infty\left(A_i(x) - V^\top D_i\right) \le D_i^\top u \qquad i = 1, \ldots, L, & \qquad (10b) \end{cases}$$

holds, then x also satisfies (P).

Proof. We know from Theorem 1 that (P) is equivalent to (7). Here, we restrict the adjustable variable λ to be a linear decision rule, that is,

$$\lambda = u + Vw.$$

Substituting this decision rule in the first constraint yields

$$\begin{aligned} & & d^\top\lambda + b(x)^\top w - f^*(w) \le 0 & \qquad \forall w \in \operatorname{dom} f^* \\ \Rightarrow & \quad d^\top u + \sup_{w\in\operatorname{dom} f^*} & \left\{\left(b(x) + V^\top d\right)^\top w - f^*(w)\right\} \le 0 & \\ \Longleftrightarrow & & d^\top u + f\left(b(x) + V^\top d\right) \le 0, & \end{aligned}$$

where the last equivalence follows from the definition of the conjugate. For the second constraint we find

$$\begin{aligned} & D_i^\top\lambda \ge A_i(x)^\top w & \qquad \forall w \in \operatorname{dom} f^*,\ i = 1, \ldots, L \\ \Rightarrow \quad & \delta^*\left(A_i(x) - V^\top D_i \,\middle|\, \operatorname{dom} f^*\right) \le D_i^\top u & i = 1, \ldots, L \\ \Longleftrightarrow \quad & f_\infty\left(A_i(x) - V^\top D_i\right) \le D_i^\top u & i = 1, \ldots, L, \end{aligned}$$

where the last equivalence follows from (2). □

In general, the tractability of the resulting problem (10) is dependent on the original function f and f_∞, where it is important to remark that f_∞ is conically representable by definition. Furthermore, we know for any f that is positively homogeneous, it holds that $f_\infty = f$ and thus (10) can be solved efficiently if and only if the original nominal problem can be. Sections 3.1 and 3.3 describe examples for which problem (10) can be solved about as efficiently as the original nominal problem, where f is not positively homogeneous.

Possible other techniques that can be used to solve or approximate the adjustable robust linear formulation given in Theorem 1 include, but are not limited to, finite adaptability, see e.g. (Bertsimas and Caramanis, 2007; Postek and Den Hertog, 2016), Fourier-Motzkin elimination (Zhen et al., 2018) and nonlinear decision rules, done by (Ben-Tal et al., 2004; Bertsimas et al., 2011).

Alternatively, a tighter safe approximation than the one described in Theorem 2 can be found by lifting the nonlinearity of $f^*(w)$ to the uncertainty set and using a slightly more involved decision rule.

Theorem 3. *If there exist $u \in \mathbb{R}^q$, $V \in \mathbb{R}^{q\times p}$ and $r \in \mathbb{R}^q$ for a given $x \in \mathbb{R}^n$ such that*

$$\begin{cases} d^\top u + \left(1 + d^\top r\right) f\left(\frac{V^\top d + b(x)}{1+d^\top r}\right) \leq 0 \\ 1 + d^\top r \geq 0 \\ -D_i^\top u + \left(-D_i^\top r\right) f\left(\frac{A_i(x) - V^\top D_i}{-D_i^\top r}\right) \leq 0 \qquad i = 1,\dots,L, \\ -D_i^\top r \geq 0 \end{cases} \tag{11a, 11b}$$

holds, then x also satisfies (P). *Here, D_i and $A_i(x)$ denote the i-th column of D and $A(x)$, respectively. Moreover,* (11) *is a tighter safe approximation than* (10).

Proof. From Theorem 1 we know (P) is equivalent to (7). We lift the nonlinear term $f^*(w)$ to the uncertainty set, that is, we introduce an auxiliary uncertain parameter w_0 such that we find that (7) is equivalent to

$$\forall \begin{pmatrix} w_0 \\ w \end{pmatrix} \in W,\ \exists \lambda \in \mathbb{R}^q : \begin{cases} d^\top \lambda + w^\top b(x) + w_0 \leq 0 & (12a) \\ D^\top \lambda \geq A(x)^\top w, & (12b) \end{cases}$$

where the new uncertainty set W is defined by

$$W = \left\{ \begin{pmatrix} w_0 \\ w \end{pmatrix} \in \mathbb{R}^{p+1} \;\middle|\; w_0 + f^*(w) \leq 0 \right\}.$$

The support function of this new uncertainty set is essential for deriving a tractable robust counterpart and is equal to:

$$\begin{aligned} \delta^*\left(\begin{pmatrix} z_0 \\ z \end{pmatrix} \middle| W\right) &= \sup_{(w_0\ w)^\top \in W} \left\{ z_0 w_0 + z^\top w \right\} \\ &= \begin{cases} \sup_{w\in\mathbb{R}^p} \left\{ z^\top w - z_0 f^*(w) \right\} & \text{if } z_0 \geq 0 \\ +\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} z_0 \sup_{w\in\mathbb{R}^p} \left\{ w^\top \frac{z}{z_0} - f^*(w) \right\} & \text{if } z_0 \geq 0 \\ +\infty & \text{otherwise} \end{cases} \\ &= \begin{cases} z_0 f\left(\frac{z}{z_0}\right) & \text{if } z_0 \geq 0 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{13}$$

Now, we once again use a linear decision rule for λ of the form

$$\lambda = u + Vw + rw_0, \tag{14}$$

where $u \in \mathbb{R}^q$, $V \in \mathbb{R}^{q\times p}$ and $r \in \mathbb{R}^q$, and thus we obtain a safe approximation for (P). Substituting this decision rule in (12a) yields

$$\begin{aligned}
& & d^\top \lambda + b(x)^\top w + w_0 \le 0 \qquad & \forall \begin{pmatrix} w_0 \\ w \end{pmatrix} \in W \\
\Rightarrow & & d^\top (u + Vw + rw_0) + b(x)^\top w + w_0 \le 0 \qquad & \forall \begin{pmatrix} w_0 \\ w \end{pmatrix} \in W \\
\iff & & d^\top u + \begin{pmatrix} w_0 \\ w \end{pmatrix}^\top \begin{pmatrix} 1 + d^\top r \\ V^\top d + b(x) \end{pmatrix} \le 0 \qquad & \forall \begin{pmatrix} w_0 \\ w \end{pmatrix} \in W \\
\iff & & d^\top u + \delta^* \left(\begin{pmatrix} 1 + d^\top r \\ V^\top d + b(x) \end{pmatrix} \middle| W \right) \le 0 & \\
\iff & & \begin{cases} ld^\top u + \left(1 + d^\top r\right)^\top f\left(\frac{V^\top d + b(x)}{1 + d^\top r}\right) \le 0 \\ 1 + d^\top r \ge 0 \end{cases}, &
\end{aligned} \tag{15}$$

where the last equivalence holds because of the definition of the support of W in (13). Note that (15) is exactly (11a). Similarly, substituting the linear decision rule for λ in (12b) we find

$$\begin{aligned}
& & D_i^\top \lambda \ge A_i(x)^\top w \qquad & \forall \begin{pmatrix} w_0 \\ w \end{pmatrix} \in W, \;\; i = 1, \ldots, L \\
\Rightarrow & & -D_i^\top u + \begin{pmatrix} w_0 \\ w \end{pmatrix}^\top \begin{pmatrix} -D_i^\top r \\ A_i(x) - V^\top D_i \end{pmatrix} \le 0 \qquad & \forall \begin{pmatrix} w_0 \\ w \end{pmatrix} \in W, \;\; i = 1, \ldots, L \\
\iff & & -D_i^\top u + \delta^* \left(\begin{pmatrix} -D_i^\top r \\ A_i(x) - V^\top D_i \end{pmatrix} \middle| W \right) \le 0 \qquad & i = 1, \ldots, L \\
\iff & & \begin{cases} -D_i^\top u + \left(-D_i^\top r\right) f\left(\frac{A_i(x) - V^\top D_i}{-D_i^\top r}\right) \le 0 \\ -D_i^\top r \ge 0 \end{cases} \qquad & i = 1, \ldots, L,
\end{aligned}$$

which is exactly (11b). Because the decision rule (14) is a more general decision rule than the decision rule used in Theorem 2, which equals (14) for $r = 0$, it follows that (11) is a tighter safe approximation than (10). □

The resulting system of inequalities (11) is convex and includes $2q + qp + n$ variables compared to the original n in (P). We note that the nonnegativity constraints introduced due to the domain of the perspective function relate to the initial uncertainty set U. In fact, by appropriately applying Farkas' Lemma, we can show that if the initial uncertainty set U is nonempty, we know that for any r such that $-1 < d^\top r \leq 0$, it holds that $-D_i^\top r \geq 0$.

Recall that Theorem 3 states that (11) is a tighter approximation than (10), which was proven by remarking that the decision rule used to derive (10) is a special case of the decision rule used to derive (11). Another insightful way to prove this relation is inserting definition (1) in (10), such that it is remarkably similar to (11). In particular, we find that (11) is a tighter formulation, as it allows freedom in choosing the perspective parameter through the variable r, whereas in (10) only the limit of the perspective at zero is considered.

2.3 Alternative Formulations

In Robust Optimization, robust counterparts of equivalent deterministic formulations are not necessarily equivalent (Gorissen et al., 2015). In the same spirit, safe approximations obtained by our approach from equivalent uncertain formulations are not necessarily equivalent. In this section, we explore alternative formulations to the initial formulation of the problem and comment on the effects of using these formulations on the quality of the obtained safe approximation.

Our first observation is that making the constraint 'more linear', that is, transforming it with a strictly increasing, concave function, can potentially result in a tighter safe approximations. The following theorem formalizes this idea.

Theorem 4. *Let X denote the range of f and let $g : X \to \mathbb{R}$ be concave, strictly increasing and differentiable and such that $g \circ f$ is convex. Moreover, let g be such that $g'(1) = 1$. Then applying Theorem 3 to*

$$g\left(f\left(A(x)\zeta + b(x)\right)\right) \leq g(1) \qquad \forall \zeta \in U, \tag{16}$$

yields a safe approximation that is at least as tight as the approximation obtained by applying Theorem 3 to

$$f\left(A(x)\zeta + b(x)\right) \leq 1 \qquad \forall \zeta \in U. \tag{17}$$

Proof. Applying Theorem 3 to (17) yields

$$\begin{cases} d^\top u + \left(1 + d^\top r\right)\left[f\left(\dfrac{V^\top d + b(x)}{1 + d^\top r}\right) - 1\right] \leq 0 & & \text{(18a)} \\ 1 + d^\top r \geq 0 & & \text{(18b)} \\ -D_i^\top u + \left(-D_i^\top r\right)\left[f\left(\dfrac{A_i(x) - V^\top D_i}{-D_i^\top r}\right) - 1\right] \leq 0 & i = 1, \dots, L & \text{(18c)} \\ -D_i^\top r \geq 0 & i = 1, \dots, L, & \text{(18d)} \end{cases}$$

as a safe approximation. If, on the other hand, we apply Theorem 3 to (16) we obtain the following safe approximation:

$$\begin{cases} d^\top u + \left(1 + d^\top r\right)\left[g\left(f\left(\dfrac{V^\top d + b(x)}{1 + d^\top r}\right)\right) - g(1)\right] \leq 0 & & \text{(19a)} \\ 1 + d^\top r \geq 0 & & \text{(19b)} \\ -D_i^\top u + \left(-D_i^\top r\right)\left[g\left(f\left(\dfrac{A_i(x) - V^\top D_i}{-D_i^\top r}\right)\right) - g(1)\right] \leq 0 & i = 1, \dots, L & \text{(19c)} \\ -D_i^\top r \geq 0 & i = 1, \dots, L, & \text{(19d)} \end{cases}$$

Observe that (18a) is equivalent to

$$f\left(\frac{V^\top d + b(x)}{1 + d^\top r}\right) \leq 1 - \frac{d^\top u}{1 + d^\top r}, \tag{20}$$

and (19a) is equivalent to

$$f\left(\frac{V^\top d + b(x)}{1 + d^\top r}\right) \leq g^{-1}\left(g(1) - \frac{d^\top u}{1 + d^\top r}\right). \tag{21}$$

Therefore, (19a) leads to a tighter safe approximation than (18a) if (20) implies (21), that is, if

$$1 - \frac{d^\top u}{1 + d^\top r} \leq g^{-1}\left(g(1) - \frac{d^\top u}{1 + d^\top r}\right) \iff g\left(1 - \frac{d^\top u}{1 + d^\top r}\right) \leq g(1) - \frac{d^\top u}{1 + d^\top r}.$$

From Taylor's theorem and the fact that g is concave, we know that

$$g\left(1 - \frac{d^\top u}{1 + d^\top r}\right) \leq g(1) + \frac{-d^\top u}{1 + d^\top r} g'(1) = g(1) + \frac{-d^\top u}{1 + d^\top r},$$

where we use that $g'(1) = 1$ in the last inequality. Similarly, we can prove that (18c) implies (19c) and thus (19) is a tighter safe approximation than (18). □

It is important to note that any concave, strictly increasing function g can be scaled such that $g'(1) = 1$, without affecting the feasible region for x in (16). In particular, this means that applying the natural logarithm or two times the square root to a constraint yields an approximation that is at least as tight, given that the resulting constraint is still convex. An example of this can be found in geometric programming which can be convexly represented as an exponential sum or as the natural logarithm of said sum. This theorem thus shows that the latter formulation yields the tighter approximation. More information on geometric programming can be found in Section 3.4.

Our next observation regards the treatment of constants in the constraint. More specifically, we note that including a constant in the function definition leads to the exact same safe approximation as leaving it on the right-hand side does. In other words, it can be shown that for any $c \in \mathbb{R}$, applying Theorem 3 to

$$f\left(A(x)\zeta + b(x)\right) \leq c \qquad \forall \zeta \in U,$$

leads to an equivalent safe approximation as applying Theorem 3 to

$$\tilde{f}\left(A(x)\zeta + b(x)\right) \leq 0 \qquad \forall \zeta \in U,$$

where $\tilde{f} = f - c$.

2.4 Adversarial Approach

Although the constraints in the safe approximation (11) resulting from Theorem 3 are in general as efficient to solve as the original constraint f, the nature of the perspective can lead to numerical issues in practice (Jung et al., 2013). In this section, we discuss an alternative solution method commonly referred to as the Adversarial Approach (Bienstock and Özbay, 2008).

The adversarial approach is an iterative procedure developed to solve static robust optimization problems of which a tractable reformulation of the robust counterpart is unknown. We suggest applying this approach to an intermediate problem encountered in the proof of Theorem 3. Specifically, we suggest applying the approach to the static robust optimization problem obtained after substituting the decision rule given in (14):

$$\forall \begin{pmatrix} w \\ w_0 \end{pmatrix} \in W: \quad \begin{cases} d^\top u + \left(V^\top d + b(x)\right)^\top w + \left(1 + d^\top r\right) w_0 \leq 0 & \text{(22a)} \\ \left(A_i(x) - V^\top D_i\right)^\top w + \left(-D_i^\top r\right) w_0 \leq D_i^\top u \qquad i = 1, \ldots, L. & \text{(22b)} \end{cases}$$

The adversarial approach solves (22) by replacing W by some finite subset $\hat{W} \subseteq W$ and subsequently finds scenario(s) $(w \;\; w_0)^\top$ for which (22a) and/or (22b) are violated. Iteratively, these scenarios are added to $\hat{W}$ and the process begins anew.

Since (22) is a linear optimization problem for a discrete uncertainty set $\hat{W}$, the first step is computationally straightforward. The tractability of the second step, on the other hand, is highly dependent on W and thus on the original function f, as it is the problem of maximizing a linear function over a set defined by convex constraints. In Section 4.1 on geometric programming, we treat an example where this second problem has a closed-form solution because of some special structure of f, and thus the adversarial approach works exceptionally well.

3 Theoretical Applications

3.1 Quadratic Programming

In this section, we consider the general uncertain quadratic constraint given by

$$\zeta^\top H(x)^\top H(x)\zeta + h(x)^\top \zeta \leq g(x) \qquad \forall \zeta \in U, \tag{23}$$

where $H : \mathbb{R}^n \to \mathbb{R}^{p\times L}$, $h : \mathbb{R}^n \to \mathbb{R}^L$ and $g : \mathbb{R}^n \to \mathbb{R}$ are affine functions and U is a polyhedron as defined in (3). Because such a quadratic constraint can also be represented as a conic quadratic constraint, our approach can potentially yield two different safe approximations. The first approximation is found by defining $f : \mathbb{R}^{p+1} \to \mathbb{R}$, $A : \mathbb{R}^n \to \mathbb{R}^{(p+1)\times L}$ and $b : \mathbb{R}^n \to \mathbb{R}^{p+1}$ by:

$$f\begin{pmatrix} z \\ z_0 \end{pmatrix} = z^\top z + z_0, \qquad A(x) = \begin{bmatrix} H(x) \\ h(x)^\top \end{bmatrix}, \qquad b(x) = \begin{bmatrix} \mathbf{0} \\ -g(x) \end{bmatrix},$$

such that (23) is equivalent to

$$f\left(A(x)\zeta + b(x)\right) \leq 0 \qquad \forall \zeta \in U.$$

Theorem 3 then gives the following safe approximation for (23):

$$\begin{cases} d^\top u + \left(1 + d^\top r\right) f\begin{pmatrix} \frac{V^\top d}{1+d^\top r} \\ \frac{V_0^\top d}{1+d^\top r} \end{pmatrix} \leq g(x) & \\ \qquad\qquad 1 + d^\top r \geq 0 & \qquad\qquad\qquad\qquad (24a) \\ \left(-D_i^\top r\right) f\begin{pmatrix} \frac{H_i(x) - V^\top D_i}{-D_i^\top r} \\ \frac{h_i(x) - V_0^\top D_i}{-D_i^\top r} \end{pmatrix} \leq D_i^\top u & \quad i = 1,\ldots,L, \\ \qquad\qquad -D_i^\top r \geq 0 & \qquad\qquad\qquad\qquad (24b) \end{cases}$$

where $u \in \mathbb{R}^q$, $V \in \mathbb{R}^{q\times p}$, $V_0 \in \mathbb{R}^q$ and $r \in \mathbb{R}^q$.

Alternatively, one can show that (23) is equivalent to:

$$\left\| \begin{matrix} \left(1 + h(x)^\top \zeta - g(x)\right)/2 \\ H(x)\zeta \end{matrix} \right\|_2 \leq \left(1 - h(x)^\top \zeta + g(x)\right)/2 \qquad \forall \zeta \in U.$$

Defining $\tilde{f} : \mathbb{R}^{p+2} \to \mathbb{R}$, $\tilde{A} : \mathbb{R}^n \to \mathbb{R}^{(p+2)\times L}$ and $\tilde{b} : \mathbb{R}^n \to \mathbb{R}^{p+2}$ by

$$\tilde{f}\begin{pmatrix} z \\ \hat{z} \\ z_0 \end{pmatrix} = \left\| \begin{matrix} z \\ \hat{z} \\ z_0 \end{matrix} \right\|_2 + z_0, \qquad \tilde{A}(x) = \begin{bmatrix} \frac{1}{2}h(x)^\top \\ \frac{1}{2}h(x)^\top \\ H(x) \end{bmatrix}, \qquad \tilde{b}(x) = \begin{bmatrix} -\frac{1+g(x)}{2} \\ \frac{1-g(x)}{2} \\ 0 \end{bmatrix},$$

we can write this equivalently as

$$\tilde{f}\left(\tilde{A}(x)\zeta + \tilde{b}(x)\right) \leq 0 \qquad \forall \zeta \in U.$$

We note that $\tilde{f}$ is positively homogeneous and thus Theorem 2 and 3 yield the same safe approximation:

$$\begin{cases} d^\top u + d^\top V_0 + \left\| \begin{matrix} V^\top d \\ d^\top r + (1-g(x))/2 \end{matrix} \right\|_2 \leq \dfrac{1+g(x)}{2} & & \text{(25a)} \\ \dfrac{h_i(x)}{2} + \left\| \begin{matrix} H_i(x) - V^\top D_i \\ \frac{h_i(x)}{2} - D_i^\top r \end{matrix} \right\|_2 \leq D_i^\top u + D_i^\top V_0 & i = 1, \dots, L, & \text{(25b)} \end{cases}$$

where $u \in \mathbb{R}^q$, $V \in \mathbb{R}^{q \times p}$, $V_0 \in \mathbb{R}^q$ and $r \in \mathbb{R}^q$.

The following result states that the two approximations above are equivalent, that is, the original form of the constraint is irrelevant here, as the resulting approximation from our approach has the exact same feasible region. The proof of the following theorem can be found in D.

Theorem 5. *The safe approximations* (24) *and* (25) *are equivalent, that is, for any* $x \in \mathbb{R}^n$ *for which there exist* $u \in \mathbb{R}^q$*,* $V \in \mathbb{R}^{q \times p}$*,* $V_0 \in \mathbb{R}^q$ *and* $r \in \mathbb{R}^q$ *that satisfy* (24)*, there exist* $u \in \mathbb{R}^q$*,* $V \in \mathbb{R}^{q \times p}$*,* $V_0 \in \mathbb{R}^q$ *and* $r \in \mathbb{R}^q$ *that satisfy* (25) *and vice versa.*

3.2 Sum-of-Max Constraints

In this section we consider constraints that are the sum of maxima of multiple arguments. Constraints of this form are often used to penalize undesirable characteristics of a solution. An example of this can be found in radiotherapy optimization, where failing to deliver the prescribed dose of radiation to the target should be penalized while exceeding the prescribed dose should not be (Shepard et al., 1999; De Boeck et al., 2014). Mathematically this can be achieved by penalizing the maximum of the difference between the prescribed dose and actual dose and zero.

We consider $f : \mathbb{R}^p \to \mathbb{R}$ to be given by

$$f(v) = \sum_{i=1}^{m} \max_{k \in K_i} \{v_k\}, \tag{26}$$

such that (P) is given by

$$\sum_{j=1}^{m} \max_{k \in K_i} \{A_k(x)\zeta + b_k(x)\} \leq 0 \qquad \forall \zeta \in U, \tag{27}$$

where $K_j \subseteq \{1, \dots, p\}$, for $i = j, \dots, m$ and $A_k(x)$ is the k-th row of $A(x)$. Because f is positively homogeneous, we know from Theorem 2 that $f_\infty = f$, and a safe approximation for (27) is thus given by

$$\begin{cases} d^\top u + f\left(b(x) + V^\top d\right) \leq 0 & & \text{(28a)} \\ f\left(A_i(x) - V^\top D_i\right) \leq D_i^\top u & i = 1, \dots, L. & \text{(28b)} \end{cases}$$

Traditionally, the robust counterpart of a sum-of-max constraint is reformulated by introducing auxiliary adjustable variables y_i to reformulate (27) (Gorissen and Den Hertog, 2013; Ardestani-Jaafari and Delage, 2016):

$$\forall \zeta \in U, \;\; \exists y \in \mathbb{R}^m : \begin{cases} \displaystyle\sum_{j=1}^{m} y_j \leq 0 & \text{(29a)} \\ A_k(x)\zeta + b_k(x) \leq y_j \qquad \forall k \in K_j, \;\; i = 1, \ldots, m. & \text{(29b)} \end{cases}$$

If linear decision rules are used to solve (29), the resulting approximation coincides with (28). For details on proving this we refer to Appendix E. Ardestani-Jaafari and Delage (2016) show that for box and budget uncertainty sets, linear decision rules are optimal in solving (29) under some additional assumptions regarding the structure of $A(x)$. By the equivalence to our approach for linear decision rules, this means that linear decision rules in our approach are also optimal for box and budget uncertainty sets under these additional assumptions.

For polyhedral uncertainty sets that are not a box or a budget uncertainty set or for which the additional assumptions made by Ardestani-Jaafari and Delage (2016) are not satsified, linear decision rules are not necessarily optimal. In this case, using the approach we suggest can be beneficial as it allows for other techniques from adjustable robust optimization to be used, such as nonlinear decision rules or Fourier-Motzkin elimination.

When considering sum-of-max constraints, we can in fact also apply our approach for ellipsoidal uncertainty, without approximating the ellipsoidal uncertainty set by a polyhedron. To accomplish this, we use the fact that $f^*(w) = 0$ on its domain and that this domain is a simplex, and thus a polyhedron. This means we can apply Theorem 1 twice to find an equivalent linear ARO problem. It turns out that, in fact, this linear ARO problem is exactly (29) with the original ellipsoidal uncertainty set.

The results in this section suggest that the techniques we propose to tackle hard uncertain convex inequalities can coincide or generalize existing linearization techniques involving adjustable variables. In particular, we find that for sum-of-max constraints we obtain a reformulation that allows for more advanced adjustable robust optimization techniques to be used than just a simple linear decision rule. While applying more advanced decision rules in (29) can be cumbersome due to U being a general polyhedron, it is much easier in the adjustable formulation we obtain, as there the uncertainty set is a cartesian product of simplices. Moreover, eliminating an adjustable variable y_j with Fourier-Motzkin elimination in (29) simply results in enumerating all possible options with regard to the j-th maximum.

3.3 Sum-of-Max Squared

A more intricate version of a sum-of-max constraint is obtained by squaring the maxima before summing them. This type of constraint or penalty function is particularly interesting for problems where heavily violating a single goal is (much) more problematic than moderately violating a couple of goals. An example of such a problem is a cancer treatment planning problem, where the homogeneity of the dose administered to the target volume is an important consideration. For ease of exposition, we focus on functions $f : \mathbb{R}^p \to \mathbb{R}$ of the form

$$f(v) = \sum_{j=1}^{p} \max\left\{v_j, 0\right\}^2.$$

To apply our approximation method, we note that

$$f_\infty(v) = \begin{cases} 0 & \text{if } v \leq 0 \\ +\infty & \text{otherwise,} \end{cases}$$

and

$$\lambda f\left(\frac{v}{\lambda}\right) = \frac{1}{\lambda} f(v),$$

that is, f is positively homogeneous of order 2. A safe approximation of the constraint

$$f\left(A(x)\zeta + b(x)\right) \leq 0 \qquad \forall \zeta \in U,$$

is thus given by applying Theorem 2:

$$\begin{cases} d^\top u + \sum_{j=1}^{p} \max\left\{b_j(x) + V_j^\top d, 0\right\}^2 \leq 0 & \\ D_i^\top u \geq 0 & i = 1, \ldots, L \\ A_i(x) - V^\top D_i \leq 0 & i = 1, \ldots, L. \end{cases}$$

A tighter safe approximation can be found by applying Theorem 3. For this particular choice of f, (11) can be reformulated to:

$$\begin{cases} d^\top u + \sum_{j=1}^{p} y_j^0 \leq 0 & \\ \left\| \begin{matrix} z_j^0 \\ \frac{1}{2}\left(1 + d^\top r - y_j^0\right) \end{matrix} \right\|_2 \leq \frac{1}{2}\left(1 + d^\top r + y_j^0\right) & j = 1, \ldots, p \\ z^0 \geq 0 & \\ z^0 \geq b(x) + V^\top d & \\ 1 + d^\top r \geq 0 & \\ \sum_{j=1}^{p} y_j^i \leq D_i^\top u & i = 1, \ldots, L \\ \left\| \begin{matrix} z_j^i \\ \frac{1}{2}\left(-D_i^\top r - y_j^i\right) \end{matrix} \right\|_2 \leq \frac{1}{2}\left(-D_i^\top r + y_j^i\right) & j = 1, \ldots, p,\ \ i = 1, \ldots, L \\ z^i \geq 0 & i = 1, \ldots, L \\ z^i \geq A_i(x) - V^\top D_i & i = 1, \ldots, L \\ -D_i^\top r \geq 0 & i = 1, \ldots, L, \end{cases}$$

where the auxiliary variables $z^0, z^i \in \mathbb{R}^p$ model the maximum of $V^\top d + b(x)$ and 0 and $A_i(x) - V^\top D_i$ and 0, for $i = 1, \ldots, L$, respectively. Furthermore, the auxiliary variables $y^0, y^i \in \mathbb{R}^p$ model $\frac{\left(z_j^0\right)^2}{1+d^\top r}$ and $\frac{\left(z_j^i\right)^2}{-D_i^\top r}$, respectively.

We demonstrate the use of the above safe approximations with a numerical example in Section 4.2.

3.4 Geometric Programming

In general, a geometric programming constraint is given by (Boyd et al., 2007):

$$f(Cx + c) \leq 0, \tag{32}$$

for some $C \in \mathbb{R}^{p \times n}$, $c \in \mathbb{R}^p$, where f is the log-sum-exp function given by

$$f(z) = \log\left(e^{z_1} + \cdots + e^{z_p}\right).$$

Note that we choose this particular formulation of a geometric programming constraint as Theorem 4 shows that this results in a potentially tighter safe approximation than using simply the sum of exponential terms. In this paper, we focus on uncertainty in C. We note that h is a convex function in both the uncertain parameters C and the decision variables x. As the argument in (32) is affine in both C and x we note that there exist affine mappings $A : \mathbb{R}^n \to \mathbb{R}^{p \times L}$ and $b : \mathbb{R}^n \to \mathbb{R}^p$ such that the robust counterpart of (32) is given by

$$f(A(x)\zeta + b(x)) \leq 0 \qquad \forall \zeta \in U, \tag{33}$$

where ζ contains all elements of C and thus $L = p \cdot n$ and U is the polyhedron defined in (3).

Recall that the safe approximation resulting from Theorem 3 involves the perspective of the log-sum-exp function. We know that a constraint modeled by f can be represented as an exponential cone (Serrano, 2015), and problems including such constraints can thus be solved efficiently. Appendix B shows that any perspective of a conically representable function is conically representable with the same cone, and thus (11) can be solved efficiently in theory as it can be represented using exponential cones.

In practice, however, interior point methods can have trouble solving problems including perspective functions as gradient computations can be difficult when the denominator is close to zero. Alternatively, we can use the safe approximation as formulated in Theorem 2. To this end we note that the recession function of f is given by:

$$f_\infty(y) = \max \{ y_i \mid i = 1, \ldots, n \},$$

and thus we find that

$$\begin{cases} d^\top u + f\left(V^\top d + b(x)\right) \leq 0 & \qquad (34a) \\ \max\limits_k \left\{ A_{ki}(x) - V_k^\top D_i \right\} \leq D_i^\top u \qquad i = 1, \ldots, L, & \qquad (34b) \end{cases}$$

is a safe approximation of (33), where $A_{ki}(x)$ is the element on the k-th row and i-th column of $A(x)$ and V_k is the k-th column of V. We note that (34b) can easily be reformulated as linear constraints.

4 Numerical Results

4.1 Geometric Programming

For our first numerical experiment we tested our approach on several randomly generated geometric programming instances. In order to compare our approach with techniques from the existing literature, we use identically structured instances to Hsiung et al. (2008). In particular, this means we treat geometric programming problems with a linear objective, and a number of two-term log-sum-exp inequality constraints with uncertainty:

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \log\left(e^{\left(-\mathbf{1} + B_i^{(1)}\zeta\right)^\top x} + e^{\left(-\mathbf{1} + B_i^{(2)}\zeta\right)^\top x} \right) \leq 0 \qquad \forall \zeta \in U,\ i = 1, \ldots, m, \end{aligned} \tag{35}$$

	Approximation Error (%)	Computation Time (s)
Lower bound	-0.00	5.2
Theorem 3	0.02	117.6
Theorem 2	1.44	3.8

Table 1: Average solution quality and computation time for 20 randomly generated instances of type (35) with $n = m = 100$ and $L = 5$.

where $c = \mathbf{1} \in \mathbb{R}^n$ is the all ones vector, and $B_i^{(1)}$, $B_i^{(2)} \in \mathbb{R}^{n\times L}$ are randomly generated sparse matrices with sparsity density 0.1 whose nonzero elements are uniformly distributed on the interval $[0, 1]$. The uncertainty set U is assumed to be a box, that is,

$$U = \left\{\zeta \in \mathbb{R}^L \;\middle|\; \|\zeta\|_\infty \leq 1\right\}.$$

Note that since U is symmetric around 0, we can restrict $B_i^{(1)}$, $B_i^{(2)}$ to be nonnegative.

We first consider a set of 20 small examples with $n = m = 100$ and $L = 5$. Since L, the number of uncertain parameters, is small, (35) can be solved exactly by enumerating the 2^L vertices of U. For larger L, however, we need to resort to comparing our solutions' objective value to a lower bound. To this end, we use a lower bound based on the work of Hadjiyiannis et al. (2011) and Zhen et al. (2017) that uses the optimal solution to a safe approximation to find potentially critical scenarios in the uncertainty set. The lower bound is then constructed by solving a model that only safeguards for this finite set of critical scenarios. For more details we refer the reader to Appendix C.

To evaluate the quality of the obtained lower bound, we have included the approximation error of the lower bound compared to the exact robust objective value in Table 1 as well as the approximation error compared to the exact robust solution and computation time of the solutions to the safe approximations resulting from Theorem 2 and 3. We define this approximation error (in percentage) equally to (Hsiung et al., 2008):

$$100\left(\frac{e^{c^\top \hat{x}}}{e^{c^\top x^*}} - 1\right),$$

where $\hat{x}$ is the solution to our safe approximation and x^* is the exact robust solution. In other words, we compare the objective value of the robust geometric programming problem in posynomial form. We note that the -0.00 we report for the lower bound means we are unable to differentiate the objective value from the optimal objective value within a reasonable numerical precision.

Clearly, for instances of this size the lower bound is particularly good. Moreover, it is an order of magnitude closer to the exact robust objective value compared to the solutions we find using our safe approximation. Therefore, we expect that using the lower bound instead of the exact robust solution for larger instances has hardly any effect on the approximation error we report.

We also note that the safe approximation resulting from Theorem 3 is particularly hard to solve. In the ensuing presented numerical results, we have therefore solved this approximation with a cutting plane method as suggested by Bienstock and Özbay (2008), described in Section 2.4.

For the instances we consider, this is particularly efficient as the problem is linear and finding a violating scenario $w \in W$ is a two-dimensional problem and can be solved analytically. In particular, the cutting plane method described above solves the problem about four times faster than using a nonlinear solver for $L = 5$. The difference in speed grows as the instance size increases, which informed our decision to solely use this cutting plane method in the results that follow.

To analyze how our approach scales with more uncertain parameters, Figure 1 shows the average difference to the lower bound and computation time of both approximations for several values of L over 20 random instances. There is a clear difference between the two decision rules in approximation quality and computation time. The more involved decision rule used in Theorem 3 performs very well, having an approximation error below 0.5% for all instances considered. The simple decision rule used in Theorem 2 performs quite a bit worse with an approximation error between 1% and 5% at first, and quickly increases for larger values of L. It is, however, extremely quick to solve and its computation time seems to hardly increase for higher values of L. The computation time needed

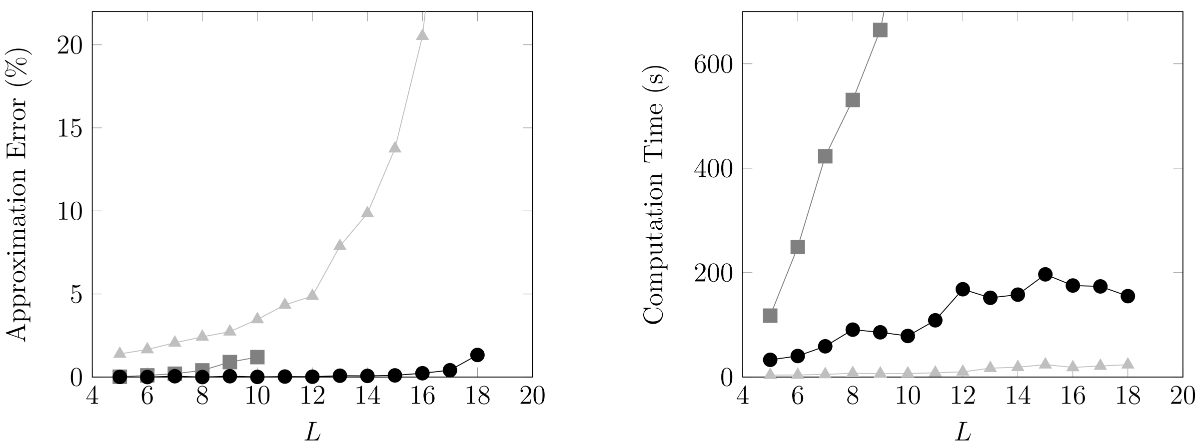

Figure 1: Average results of solving approximations to (35) over twenty randomly generated instances. Results corresponding to Theorem 2 are depicted in green triangles, results corresponding to Theorem 3 are shown as blue squares and the results from applying the cutting plane approach are depicted as orange circles.

for the complex decision rule, on the other hand, is a fair bit higher and does clearly increase as L does. We note that no results for $L = 19$ or $L = 20$ are included, as a large proportion of the randomly generated instances of this size were infeasible.

We note that Hsiung et al. (2008) report approximation errors between 30% and 0.1% dependent on the quality of approximation used, for $L = 5$ and $n = m = 500$.

The solutions to the geometric programming problems have been obtained using Julia with the JuMP interface (Dunning et al., 2017) and the IPOPT solver. The experiments were conducted on a desktop with 8 GB RAM and a 3.4 GHz Intel Core i7 processor.

4.2 Radiotherapy Optimization

Our second numerical experiment concerns a specific problem from radiotherapy optimization: inverse treatment planning of beam-on times for 3D small animal radiotherapy (Balvert et al., 2015). The core problem in treatment planning is ensuring a sufficient dose γ of radiation to the planning target volume (PTV) while minimizing the dose to the tissue around that target volume, also known as the organs at risk ($\mathcal{OAR}$). To this end, we are interested in minimizing a weighted combination of the dose 'shortage' in the PTV and the dose delivered to the $\mathcal{OAR}$. The decision variables in this problem are the locations and beam-on times for all beams used. In this specific application, we assume the beam locations are given and we attempt to find optimal beam-on times t.

It is customary in radiotherapy optimization to discretize each tissue structure into voxels. Sets of these voxels are denoted by $\mathcal{I}_{PTV}$ and $\mathcal{I}_s$ for all $s \in \mathcal{OAR}$, respectively. The dose delivered to a tissue structure is then computed as the average dose delivered to its voxels. Given these voxels, one can compute the dose rates from all beams to all voxels, referred to as the matrix Γ. The i-th row of this matrix, Γ_i, then corresponds to the dose rate of all beams to voxel i. We specifically consider the following mathematical optimization problem:

$$\begin{aligned}
\min_{t,\tau} \quad & \tau && \text{(36a)}\\
\text{s.t.} \quad & w_{PTV} \frac{1}{|\mathcal{I}_{PTV}|} \sum_{i \in \mathcal{I}_{PTV}} \max\left\{\gamma - \Gamma_i^\top t, 0\right\}^2 \\
& \qquad + (1 - w_{PTV}) \sum_{s \in \mathcal{OAR}} w_s \frac{1}{|\mathcal{I}_s|} \sum_{i \in \mathcal{I}_s} \Gamma_i^\top t \le \tau && \text{(36b)}\\
& t \ge 0, && \text{(36c)}
\end{aligned}$$

which is a slight adaptation of the problem described by Balvert et al. (2015). Here, w_{PTV} and w_s for all $s \in \mathcal{OAR}$ represent predefined weights. In particular, we choose to use a squared penalty function for undelivered dose to the PTV, similar to Fredriksson (2013, Eq. 1). Irregardless of whether the regular or squared penalty function is used, little research has been done on robust or uncertain versions of (36). An important reason for this is the general convex nature of constraint (36b), along with the fact that a natural type of uncertainty in this problem is implementation error (Van Dye et al., 2013; Van der Merwe et al., 2017), which always leads to constraints that are convex in the uncertain parameters.

In this numerical example, we therefore focus on implementation error. In particular, we consider multiplicative implementation error, that is, we replace t by $t \circ (\mathbf{1} + \epsilon)$, where $\circ$ denotes the element-wise multiplication of two vectors, and ϵ is the uncertain vector that models the implementation error. We note that, at least in this context, additive implementation error of the form $t + \Delta t$ would make little sense, as this would presume that there would also potentially be some implementation error if one chooses not to use a certain beam ($t_b = 0$).

We solve both approximations we derive in this work for Case 3 discussed by Balvert et al. (2015). In this case, there are 6 different beam angles, that is, $t \in \mathbb{R}^6$, and the PTV consists of 112,738 voxels, while the four organs at risk consist of 207,974, 2,261,739, 177,165 and 212,864 voxels. We consider box uncertainty for ϵ, with three different maximum values: 0.01, 0.05 and 0.1. For all three, the solutions to the two approximations coincide. Unfortunately, due to the size of the problem (the PTV consisted of over 100,000 voxels), we cannot obtain the exact robust solution. We are, however, able to obtain lower bounds using the technique described in Appendix C and find that the approximation error is bounded from above by 1.13%, 5.57% and 10.91% for a maximum implementation error of 1%, 5% and 10%, respectively. Furthermore, we find that the nominal solution performs 4.4% worse in the worst-case than the robust solution we find, which in turn performs 4.9% worse than the nominal solution when no uncertainty is present. It should be noted that the approximation resulting from Theorem 2 can be solved in a matter of seconds, much like the model without uncertainty.

All results in this section have been obtained using Julia with the JuMP interface (Dunning et al., 2017) and the Gurobi solver. The experiments were conducted on a desktop with 8 GB RAM and a 3.4 GHz Intel Core i7 processor.

5 Conclusions

In Robust Optimization, finding a tractable reformulation of the robust counterpart of the uncertain inequalities of interest is essential. While a systematic approach to find such tractable reformulations already exists when the inequalities of interest are concave in the uncertain parameters, no general results are available when they are convex. This paper fills that gap for polyhedral uncertainty sets by providing a reformulation of such problems to linear adjustable robust optimization problems. These problems can be approximately solved by a variety of techniques available in the literature.

An evident possibility for future research is extending this work to other types of uncertainty sets, such as ellipsoidal uncertainty sets or more general convex sets. We do remark that for ellipsoidal uncertainty sets, polyhedral outer approximations from (Ben-Tal and Nemirovski, 2001) could be used to enable the use of the theorems derived in this paper. Note that the ARO problem obtained in Theorem 1 then no longer yields an equivalent problem, but a safe approximation instead of the original problem with ellipsoidal uncertainty. Other non-polyhedral sets could potentially be treated similarly, that is, first be approximated by a polyhedron such that the results in this paper apply. As uncertainty sets are typically user specified, one could adjust their choice such that the resulting approximation is not overly conservative.

Acknowledgments

We thank Erick Delage from HEC Montréal for sharing the proof to Theorem 1, which is more concise than our original proof. The research of the first author was funded by the Netherlands Organisation for Scientific Research (NWO) Research Talent [Grant 406.17.511].

References

Alfandari, L. and García, J. C. E. (2018). Robust optimization for non-linear impact of data variation. *Computers & Operations Research*.

Ardestani-Jaafari, A. and Delage, E. (2016). Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494.

Auslender, A. and Teboulle, M. (2006). *Asymptotic cones and functions in optimization and variational inequalities*. Springer Science & Business Media.

Balvert, M., van Hoof, S. J., Granton, P. V., Trani, D., den Hertog, D., Hoffmann, A. L., and Verhaegen, F. (2015). A framework for inverse planning of beam-on times for 3d small animal radiotherapy using interactive multi-objective optimisation. *Physics in Medicine & Biology*, 60(14):5681–5698.

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

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization*. Princeton University Press.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376.

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

Ben-Tal, A. and Nemirovski, A. (2001). On polyhedral approximations of the second-order cone. *Mathematics of Operations Research*, 26(2):193–205.

Ben-Tal, A., Nemirovski, A., and Roos, C. (2002). Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560.

Bertsimas, D. and Caramanis, C. (2007). Adaptability via sampling. In *2007 46th IEEE Conference on Decision and Control*, pages 4717–4722. IEEE.

Bertsimas, D., Iancu, D. A., and Parrilo, P. A. (2011). A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12):2809–2824.

Bertsimas, D. and Sim, M. (2006). Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1-2):5–36.

Bienstock, D. and Özbay, N. (2008). Computing robust basestock levels. *Discrete Optimization*, 5(2):389–414.

Boyd, S., Kim, S.-J., Vandenberghe, L., and Hassibi, A. (2007). A tutorial on geometric programming. *Optimization and Engineering*, 8(1):67–127.

Chassein, A. and Goerigk, M. (2019). On the complexity of robust geometric programming with polyhedral uncertainty. *Operations Research Letters*, 47(1):21–24.

De Boeck, L., Beliën, J., and Egyed, W. (2014). Dose optimization in high-dose-rate brachytherapy: a literature review of quantitative models from 1990 to 2010. *Operations Research for Health Care*, 3(2):80–90.

Dunning, I., Huchette, J., and Lubin, M. (2017). Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320.

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

El Ghaoui, L., Oustry, F., and Lebret, H. (1998). Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52.

Fredriksson, A. (2013). *Robust optimization of radiation therapy accounting for geometric uncertainty*. PhD thesis, KTH Royal Institute of Technology.

Gorissen, B. L. and Den Hertog, D. (2013). Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research*, 227(1):30–43.

Gorissen, B. L., Den Hertog, D., and Hoffmann, A. L. (2013). Mixed integer programming improves comprehensibility and plan quality in inverse optimization of prostate hdr brachytherapy. *Physics in Medicine & Biology*, 58(4):1041.

Gorissen, B. L., Yanıkoğlu, İ., and den Hertog, D. (2015). A practical guide to robust optimization. *Omega*, 53:124–137.

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

Hsiung, K.-L., Kim, S.-J., and Boyd, S. (2008). Tractable approximate robust geometric programming. *Optimization and Engineering*, 9(2):95–118.

Jung, M. N., Kirches, C., and Sager, S. (2013). On perspective functions and vanishing constraints in mixed-integer nonlinear optimal control. In *Facets of Combinatorial Optimization*, pages 387–417. Springer.

Lobo, M. S., Vandenberghe, L., Boyd, S., and Lebret, H. (1998). Applications of second-order cone programming. *Linear algebra and its applications*, 284(1-3):193–228.

Postek, K. and Den Hertog, D. (2016). Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS Journal on Computing*, 28(3):553–574.

Rockafellar, R. T. (1970). *Convex analysis.* Princeton University Press, Princeton, NJ.

Roos, E., den Hertog, D., Ben-Tal, A., de Ruiter, F., and Zhen, J. (2018). Approximation of hard uncertain convex inequalities. *Available on Optimization Online.*

Serrano, S. A. (2015). *Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone.* PhD thesis, Stanford University.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory*, volume 16. SIAM.

Shepard, D. M., Ferris, M. C., Olivera, G. H., and Mackie, T. R. (1999). Optimizing the delivery of radiation therapy to cancer patients. *SIAM Review*, 41(4):721–744.

Van der Merwe, D., Van Dyk, J., Healy, B., Zubizarreta, E., Izewska, J., Mijnheer, B., and Meghzifene, A. (2017). Accuracy requirements and uncertainties in radiotherapy: a report of the international atomic energy agency. *Acta Oncologica*, 56(1):1–6.

Van Dye, J., Batista, J., and Bauman, G. S. (2013). Accuracy and uncertainty considerations in modern radiation oncology. *The Modern Technology of Radiation Oncology*, 3:361–412.

Wiesemann, W., Kuhn, D., and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376.

Zhen, J., De Ruiter, F. J., and Den Hertog, D. (2017). Robust optimization for models with uncertain SOC and SDP constraints. *Optimization Online.*

Zhen, J. and Den Hertog, D. (2017). Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS Journal on Computing*, 30(1):31–42.

Zhen, J., Den Hertog, D., and Sim, M. (2018). Adjustable robust optimization via fourier–motzkin elimination. *Operations Research*, 66(4):1086–1100.

A Recession Functions

As discussed earlier, the recession function can be defined in multiple ways. In this paper, we mainly use it to concisely denote the support function of the domain of a function's conjugate. An advantage of the recession function besides concise notation is the relative ease of computing a recession function. Let $f^1, \ldots, f^m$ be convex, proper and lower semicontinuous functions. Then, the following composition rules for recession functions are valid (Auslender and Teboulle, 2006, Proposition 2.6.1, 2.6.2):

1. Let f be defined by $f(x) = \sum_{i=1}^m f^i(x)$. Then $f_\infty(y) = \sum_{i=1}^m f^i_\infty(y)$;
2. Let f be defined by $f(x) = \sup_{i\in\{1,\ldots,m\}} f^i$. Then $f_\infty(y) = \sup_{i\in\{1,\ldots,m\}} f^i_\infty(y)$.

Moreover, if $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a proper convex function, A is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $\psi : (-\infty, b) \to \mathbb{R})$ for $0 \leq b \leq +\infty$ is convex and nondecreasing with $\psi_\infty(1) > 0$ it holds that (Auslender and Teboulle, 2006, Proposition 2.6.3, 2.6.4):

3. Let f be defined by $f(x) = g(Ax)$. Then $f_\infty(y) = g_\infty(Ay)$;
4. Let f be defined by
$$f(x) = \begin{cases} \psi(g(x)) & \text{if } x \in \mathrm{dom}\,(g) \\ +\infty & \text{otherwise.} \end{cases}$$
Then
$$f_\infty(y) = \begin{cases} \psi_\infty(f_\infty(y)) & \text{if } y \in \mathrm{dom}\,(f_\infty) \\ +\infty & \text{otherwise.} \end{cases}$$

Using the above composition rules as well as the recession functions of some often encountered basic functions f, one can directly find the recession function of the function of interest. An overview of some common recession functions is given in Table 2. It should be noted that the recession function is always conically representable, as its epigraph is the recession cone of the epigraph of f and thus is a cone by definition (Rockafellar, 1970, p. 66). We additionally remark

$f(x)$	$f_\infty(y)$
$\sqrt{1 + x^\top Qx}$ $(Q \succeq 0)$	$\sqrt{y^\top Qy}$
$x^\top Qx + q^\top x + c$ $(Q \succeq 0)$	$\begin{cases} q^\top y & \text{if } Qy = 0 \\ +\infty & \text{if } Qy \neq 0 \end{cases}$
$\log \sum_{i=1}^n e^{x_i}$ $(n > 1)$	$\max\{y_i \mid i = 1, \ldots, n\}$
$\sum_{i=1}^n \sqrt{1 + x_i^2}$	$\|y\|_2$
$\sum_{i=1}^m \max_{k\in K_i} \{x_k\}$	$\sum_{i=1}^m \max_{k\in K_i} \{y_k\}$
$\|x\|_2$	$\|y\|_2$

Table 2: Some examples of functions f with recession functions f_∞.

that for all positively homogeneous functions of order one, or equivalently all functions such that $f^*(y) = 0$ on its domain, it holds that $f_\infty(x) = f(x)$.

Lemma 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a closed, convex function. It then holds that f is positively homogeneous if and only if $f^*(y) = 0$ for all $y \in \operatorname{dom}(f^*)$.*

Proof. '$\Rightarrow$': Assume f is positively homogeneous. Suppose $\exists x^* \in \operatorname{dom}(f^*)$ such that $f^*(x^*) = \alpha \neq 0$. We first consider the case where $\alpha > 0$. We know

$$\alpha = \sup_{y \in \mathbb{R}^n} \left\{ y^\top x - f(y) \right\},$$

that is, there exists a sequence $(y_k)_{k=0}^\infty$ with $y_k \in \mathbb{R}^n \quad \forall k$ such that

$$\lim_{k \to \infty} \left\{ y_k^\top x - f(y_k) \right\} = \alpha.$$

Now let $\lambda > 0$ and define the sequence $(z_k)_{k=0}^\infty$ by $z_k = \lambda y_k$. Then

$$\begin{aligned} \lim_{k \to \infty} \left\{ z_k^\top x - f(z_k) \right\} &= \lim_{k \to \infty} \left\{ \lambda y_k^\top x - f(\lambda y_k) \right\} \\ &= \lambda \lim_{k \to \infty} \left\{ y_k^\top x - f(y_k) \right\} \\ &= \lambda \alpha > \alpha, \end{aligned}$$

which is a contradiction with α being the supremum as defined by (A).

Now consider the case where $\alpha < 0$. We know for sure that

$$\alpha \geq 0^\top x - f(0) = -f(0).$$

Let $\lambda > 0$. Then, because f is positively homogeneous, we know that

$$f(0) = f(\lambda \cdot 0) = \lambda f(0),$$

and thus $f(0) = 0$. This implies $\alpha \geq 0$, which is a contradiction.

'$\Leftarrow$': Assume $f^*(y) = 0$ for all $y \in \operatorname{dom}(f^*)$. Because f is closed and convex we know

$$f(x) = f^{**}(x) = \sup_{y \in \operatorname{dom}(f^*)} \left\{ y^\top x - f^*(x) \right\} = \sup_{y \in \operatorname{dom}(f^*)} \left\{ y^\top x \right\}.$$

Let $\lambda > 0$. We find

$$\begin{aligned} f(\lambda x) &= f^{**}(\lambda x) \\ &= \sup_{y \in \operatorname{dom}(f^*)} \left\{ y^\top (\lambda x) \right\} \\ &= \lambda \sup_{y \in \operatorname{dom}(f^*)} \left\{ y^\top x \right\} \\ &= \lambda f^{**}(x) = f(x), \end{aligned}$$

and thus we find that f is positively homogeneous. $\square$

Lemma 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a closed, convex and positively homogeneous function. It then holds that $f_\infty = f$.*

Proof. We use definition (1) for the recession function to find:

$$\begin{aligned} f_\infty(x) &= \lim_{\lambda \downarrow 0} \lambda f\left(\frac{x}{\lambda}\right) \\ &= \lim_{\lambda \downarrow 0} f(x) \\ &= f(x), \end{aligned}$$

because f is positively homogeneous. □

B Proof of Conically Representable Perspective

We use the definition of conically representable from Serrano (2015), that is, a function $f : \mathbb{R}^n \to \mathbb{R}$ is conically representable if its epigraph can be written as

$$\begin{aligned} \operatorname{Epi} f = &\ \{(x,t) \mid f(x) \le t\} \\ = &\ \{(x,t) \mid \exists u \in \mathbb{R}^m,\ S(x,u,t) = 0,\ T(x,u,t) \in \mathcal{K}\}, \end{aligned}$$

where $S : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{k_1}$ and $T : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^{k_2}$ are affine mappings and $\mathcal{K}$ is a cone.

Theorem 6. *If f is conically representable, so is its perspective (fv).*

Proof. Let S, T be the affine mappings that define the conic representation of f and let $\mathcal{K}$ be the corresponding cone. Define $S^{per} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}^{k_1}$ and $T^{per} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}_{++} \to \mathbb{R}^{k_2}$ by

$$S^{per}(x,u,t,v) = vS\left(\frac{x}{v}, \frac{u}{v}, \frac{t}{v}\right), \qquad T^{per}(x,u,t,v) = vT\left(\frac{x}{v}, \frac{u}{v}, \frac{t}{v}\right).$$

Clearly, S^{per} and T^{per} are affine mappings. Moreover we find

$$\begin{aligned} \operatorname{Epi}(fv) = &\ \left\{(x,v,t) \;\middle|\; vf\left(\frac{x}{v}\right) \le t\right\} \\ = &\ \left\{(x,v,t) \;\middle|\; \left(\frac{x}{v}, \frac{t}{v}\right) \in \operatorname{Epi} f\right\} \\ = &\ \left\{(x,v,t) \;\middle|\; \exists u \in \mathbb{R}^m,\ S\left(\frac{x}{v}, u, \frac{t}{v}\right) = 0,\ T\left(\frac{x}{v}, u, \frac{t}{v}\right) \in \mathcal{K}\right\} \\ = &\ \left\{(x,v,t) \;\middle|\; \exists u \in \mathbb{R}^m,\ S\left(\frac{x}{v}, \frac{u}{v}, \frac{t}{v}\right) = 0,\ T\left(\frac{x}{v}, \frac{u}{v}, \frac{t}{v}\right) \in \mathcal{K}\right\} \\ = &\ \{(x,v,t) \mid \exists u \in \mathbb{R}^m,\ S^{per}(x,u,t,v) = 0,\ T^{per}(x,u,t,v) \in \mathcal{K}\}, \end{aligned}$$

which concludes the proof. □

C Progressive Approximation

As all sets of constraints described in Section 2.2 are safe approximations to our original constraint (P), they can yield suboptimal solutions. In particular, we propose linear decision rules to solve (7), which is equivalent to (P), of which we know they generally do not guarantee to solve adjustable robust optimization problems to optimality (Ben-Tal et al., 2004). Moreover, as our adjustable formulation (7) exhibits left-hand side uncertainty, that is, the uncertain parameter w directly interacts with decision variables x, little is known with regard to the approximative power of linear decision rules.

In this section, therefore, we focus on finding a good progressive approximation to (P) such that we can gauge the quality of the conservative approximations we propose. A simple method detailed by Zhen et al. (2017) to obtain such approximation is to only require (P) to hold for a finite subset of scenarios from the uncertainty set U. The approximation is then given by

$$f\left(A(x)\zeta^{(k)} + b(x)\right) \leq 0 \qquad k = 1, \ldots, K, \tag{37}$$

where $\left\{\zeta^{(1)}, \ldots, \zeta^{(K)}\right\} \subseteq U$. We note that these constraints are exactly as computationally tractable as the original constraint without uncertainty. In fact, because we assume a polyhedral set U and f is convex, (37) is equivalent to (P) if $\left\{\zeta^{(1)}, \ldots, \zeta^{(K)}\right\}$ contains all extreme points of U. Generally, U has prohibitively many extreme points though, and we must resort to some other way of finding scenarios $\zeta^{(1)}, \ldots, \zeta^{(K)}$.

We can apply the same reasoning as above to (7) to find an approximation:

$$\text{For } k = 1, \ldots, K, \quad \exists \lambda^{(k)} \in \mathbb{R}^q : \left\{ w_0^{(k)} + b(x)^\top w^{(k)} + d^\top \lambda^{(k)} \leq 0 \quad D^\top \lambda^{(k)} \geq A(x)^\top w^{(k)}, \right.$$

where $\left\{ \begin{pmatrix} w_0^{(1)} \\ w^{(1)} \end{pmatrix}, \ldots, \begin{pmatrix} w_0^{(K)} \\ w^{(K)} \end{pmatrix} \right\} \subset W$ and $\lambda^{(k)} \in \mathbb{R}^q$ is a non-adjustable variable. Recall that

$$W = \left\{ \begin{pmatrix} w_0 \\ w \end{pmatrix} : w_0 + f^*(w) \leq 0 \right\},$$

which generally has infinitely many extreme points.

An approach to find a small and efficient set of scenarios for two-stage fixed-recourse robust constraints is suggested by Hadjiyiannis et al. (2011). For any feasible solution $\hat{x}$ and linear decision rule $\hat{\lambda} = \hat{u} + \hat{V}w + \hat{r}w_0$, we find scenarios that are worst-case for the constraints in (7). We then hope that these scenarios are also worst-case for the actual optimal solution x^*, λ^* of (7). For our problem, this means that we obtain scenarios

$$\begin{pmatrix} \bar{w}_0 \\ \bar{w} \end{pmatrix} = \underset{\begin{pmatrix} w_0 \\ w \end{pmatrix} \in W}{\arg\max} \left\{ d^\top \left(\hat{u} + \hat{V}w + \hat{r}w_0 \right) + b(\hat{x})^\top w + w_0 \right\},$$

as well as the worst-case scenarios from (7b). An extension proposed by Zhen et al. (2017) is to use these $L+1$ scenarios to also obtain scenarios $\zeta^{(1)}, \ldots, \zeta^{(L+1)}$ by solving

$$\bar{\zeta}^{(k)} = \underset{\zeta \in U}{\arg\max} \left\{ \left(A(\hat{x})\zeta + b(\hat{x}) \right)^\top \bar{w}^{(k)} \right\}.$$

We note that similarly to this approach, we can also obtain worst-case scenarios $\bar{w}$ based on a linear decision rule solving (10). For more details, we refer to the papers by Hadjiyiannis et al. (2011) and Zhen et al. (2017).

D Proof of Equivalence of Quadratic and Conic Quadratic Approximations

Proof of Theorem 5. Let $x \in \mathbb{R}^n$ be such that there exist $u \in \mathbb{R}^q$, $V \in \mathbb{R}^{q\times p}$ and $r \in \mathbb{R}^q$ that satisfy (24). We claim choosing $u^* = \frac{1}{2}(u-r)$, $V_0^* = \frac{1}{2}V_0$ and $r^* = \frac{1}{2}(u+V_0+r)$ means that (x, u^*, V, V_0^*, r^*) satisfies (25). First suppose $1 + d^\top r > 0$. Then we find from (24a):

$$
\begin{aligned}
& d^\top u + \left(1+d^\top r\right)\left[\frac{d^\top V V^\top d}{(1+d^\top r)^2} + \frac{d^\top V_0}{1+d^\top r}\right] \le g(x) \\
\iff\quad & d^\top u + d^\top V_0 + \frac{d^\top V V^\top d}{1+d^\top r} \le g(x) \\
\iff\quad & d^\top V V^\top d + \left(1+d^\top r\right)\left(d^\top (u+V_0) - g(x)\right) \le 0 \\
\iff\quad & d^\top V V^\top d + \left[\frac{1}{2}\left(1+d^\top r + d^\top (u+V_0) - g(x)\right)\right]^2 \\
& \qquad - \left[\frac{1}{2}\left(1+d^\top r - d^\top (u+V_0) + g(x)\right)\right]^2 \le 0 \\
\iff\quad & d^\top V V^\top d + \left[\frac{1}{2}\left(1+d^\top r + d^\top (u+V_0) - g(x)\right)\right]^2 \\
& \qquad \le \left[\frac{1}{2}\left(1+d^\top r - d^\top (u+V_0) + g(x)\right)\right]^2 \\
\iff\quad & \left\| \begin{matrix} V^\top d \\ \frac{1}{2}\left(d^\top (u+V_0) + d^\top r + 1 - g(x)\right) \end{matrix} \right\|_2 \le \frac{1}{2}\left(d^\top r - d^\top (u+V_0)\right) + \frac{1+g(x)}{2} \qquad (39) \\
\iff\quad & d^\top \left(\frac{1}{2}(u+V_0-r)\right) + \left\| \begin{matrix} V^\top d \\ d^\top \left(\frac{1}{2}(u+V_0+r)\right) + \frac{1-g(x)}{2} \end{matrix} \right\|_2 \le \frac{1+g(x)}{2} \\
\iff\quad & d^\top u^* + d^\top V_0^* + \left\| \begin{matrix} V^\top d \\ d^\top r^* + \frac{1-g(x)}{2} \end{matrix} \right\|_2 \le \frac{1+g(x)}{2},
\end{aligned}
$$

where we use the fact that $\frac{1}{2}\left(1+d^\top r - d^\top (u+V_0) + g(x)\right) > 0$ for the equivalence in (39). This can be seen directly from (24a) and the realization that $d^\top V V^\top d \ge 0$ because it is a square. Suppose, on the other hand, that $1 + d^\top r = 0$. Then by our definition of the perspective function (24a) is equivalent to

$$
d^\top u + f_\infty \begin{pmatrix} V^\top d \\ V_0^\top d \end{pmatrix} \le g(x) \iff \begin{cases} d^\top u + d^\top V_0 \le g(x) \\ \qquad\qquad V^\top d = \mathbf{0}. \end{cases}
$$

Using this we find:

$$\begin{aligned}
d^\top u^* + d^\top V_0^* + \left\| \begin{matrix} V^\top d \\ d^\top r^* + \frac{1-g(x)}{2} \end{matrix} \right\|_2 &= \frac{1}{2} d^\top (u-r) + \frac{1}{2} d^\top V_0 + \frac{1}{2} d^\top (u + V_0 + r) + \frac{1-g(x)}{2} \\
&= d^\top u + d^\top V_0 + \frac{1-g(x)}{2} \\
&\le g(x) + \frac{1-g(x)}{2} \\
&= \frac{1+g(x)}{2}.
\end{aligned}$$

We have thus shown that (x, u^*, V, V_0^*, r^*) satisfies (25a).

Similarly, we can show for any $i = 1, \ldots, L$ that if $-D_i^\top r > 0$ it holds that

$$\begin{aligned}
& \left(-D_i^\top r\right) \left[\frac{\left(H_i(x) - V^\top D_i\right)^\top \left(H_i(x) - V^\top D_i\right)}{\left(-D_i^\top r\right)^2} + \frac{h_i(x) - D_i^\top V_0}{-D_i^\top r} \right] \le D_i^\top u \\
\iff\ & \frac{\left(H_i(x) - V^\top D_i\right)^\top \left(H_i(x) - V^\top D_i\right)}{-D_i^\top r} + h_i(x) - D_i^\top u - D_i^\top V_0 \le 0 \\
\iff\ & \left(H_i(x) - V^\top D_i\right)^\top \left(H_i(x) - V^\top D_i\right) + \left(1 + d^\top r\right) \left(h_i(x) - D_i^\top u - D_i^\top V_0\right) \le 0 \\
\iff\ & \left(H_i(x) - V^\top D_i\right)^\top \left(H_i(x) - V^\top D_i\right) + \left[\frac{1}{2}\left(-D_i^\top r + h_i(x) - D_i^\top u - D_i^\top V_0\right)\right]^2 \\
& \qquad - \left[\frac{1}{2}\left(-D_i^\top r - h_i(x) + D_i^\top u + D_i^\top V_0\right)\right]^2 \le 0 \\
\iff\ & \left(H_i(x) - V^\top D_i\right)^\top \left(H_i(x) - V^\top D_i\right) + \left[\frac{1}{2}\left(-D_i^\top r + h_i(x) - D_i^\top u - D_i^\top V_0\right)\right]^2 \\
& \qquad \le \left[\frac{1}{2}\left(-D_i^\top r - h_i(x) + D_i^\top u + D_i^\top V_0\right)\right]^2 \\
\iff\ & \left\| \begin{matrix} H_i(x) - V^\top D_i \\ \frac{1}{2}\left(-D_i^\top u - D_i^\top V_0 - D_i^\top r + h_i(x)\right) \end{matrix} \right\|_2 \le \frac{1}{2}\left(D_i^\top u + D_i^\top V_0 - D_i^\top r\right) - \frac{1}{2} h_i(x) \\
\iff\ & \frac{h_i(x)}{2} + \left\| \begin{matrix} H_i(x) - V^\top D_i \\ \frac{h_i(x)}{2} - D_i^\top \left(\frac{1}{2}(u + V_0 + r)\right) \end{matrix} \right\|_2 \le D_i^\top \left(\frac{1}{2}(u-r)\right) + D_i^\top \left(\frac{1}{2} V_0\right) \\
\iff\ & \frac{h_i(x)}{2} + \left\| \begin{matrix} \frac{h_i(x)}{2} - D_i^\top r^* \\ H_i(x) - V^\top D_i \end{matrix} \right\|_2 \le D_i^\top u^* + D_i^\top V_0^*.
\end{aligned}$$

Moreover, if $-D_i^\top r = 0$, it holds that

$$f_\infty \begin{pmatrix} H_i(x) - V^\top D_i \\ h_i(x) - D_i^\top V_0 \end{pmatrix} \le D_i^\top u \iff \begin{cases} h_i(x) - D_i^\top V_0 \le D_i^\top u \\ H_i(x) - V^\top D_i = \mathbf{0}, \end{cases}$$

from which it also follows that (x, u^*, V, V_0^*, r^*) satisfies (25b).

Now, let $x \in \mathbb{R}^n$ such that there exist $u^* \in \mathbb{R}^q$, $V \in \mathbb{R}^{q \times p}$, $V_0^* \in \mathbb{R}^q$ and $r^* \in \mathbb{R}^q$ that satisfy (25). We claim choosing $u = u^* + r^* - V_0^*$, $V_0 = 2V_0^*$ and $r = r^* - u^* - V_0^*$ means that (x, u, V, V_0, r) satisfies (24).

First of all, note that we defined u, V_0 and r exactly such that they are the inverse of the equations used before for u^*, V_0^* and r^*. As all steps in the derivation above are two-way implications, it directly follows that (x, u, V, V_0, r) satisfies (24).

□

E Equivalence of Sum-of-Max Reformulations

Substituting linear decision rules of the form $y_i(\zeta) = u + V\zeta$ in (29), yields the following robust counterpart:

$$\begin{cases} d^\top z_0 + \sum_{i=1}^m u_i \leq 0 & & \text{(40a)} \\ D^\top z_0 \geq \sum_{i=1}^m V_i^\top & & \text{(40b)} \\ d^\top z_{ik} - u_i + b_k(x) \leq 0 & \forall k \in K_i,\ \ i = 1, \dots, m & \text{(40c)} \\ D^\top z_{ik} \geq A_k(x)^\top - V_i^\top & \forall k \in K_i,\ \ i = 1, \dots, m & \text{(40d)} \\ z_{ik} \geq 0 & \forall k \in K_i,\ \ i = 1, \dots, m & \text{(40e)} \\ z_0 \geq 0. & & \text{(40f)} \end{cases}$$

Here, $z_0 \in \mathbb{R}^q_+$ is the dual variable corresponding to the robust counterpart of (29a) and $z_{ik} \in \mathbb{R}^q_+$ are the dual variables corresponding to the robust counterpart of (29b). Moreover, $A_k(x)$ is the k-th row of $A(x)$ and V_i is the i-th row of V in this notation.

If we continue, on the other hand, from (10), with f defined by (26), and introducing auxiliary variables z_{ij} for $i = 1, \dots, m$, $j = 0, \dots, L$, to model the sum of maxima, we obtain the following formulation:

$$\begin{cases} d^\top u + \sum_{i=1}^m z_{i0} \leq 0 & & \text{(41a)} \\ D^\top u \geq \sum_{i=1}^m z_i & & \text{(41b)} \\ d^\top V_k - z_{i0} + b_k(x) \leq 0 & \forall k \in K_i,\ \ i = 1, \dots, m & \text{(41c)} \\ D^\top V_k \geq A_k(x)^\top - z_i & \forall k \in K_i,\ \ i = 1, \dots, m. & \text{(41d)} \end{cases}$$

Here, $z_i = \begin{bmatrix} z_{i1} & \dots & z_{iL} \end{bmatrix}^\top$ and $A_k(x)$ is the k-th row of $A(x)$, while V_k is the k-th column of V.

Although the interpretation and naming of variables differs in the safe approximations (40) and (41), there is only one true difference. In (40) $z_{ik} \in \mathbb{R}^q$ for all $k \in K_i$, $i = 1, \dots, m$, while the corresponding variable $V \in \mathbb{R}^{q \times p}$ does not necessarily have the same dimension. The problems are thus only equivalent when all K_i are pair-wise disjoint, such that $\sum_{i=1}^m |K_i| = p$. We can, without loss of generality, formulate any sum-of-max constraint such that this holds, by appropriately defining $A(x)$ and $b(x)$. This implies that the two approaches are in fact equivalent.