Robust nonlinear optimization via the dual

Bram L. Gorissen[1] and Dick den Hertog[2]

Abstract

Robust nonlinear optimization is not as well developed as the linear case, and limited in the constraints and uncertainty sets it can handle. In this work we extend the scope of robust optimization by showing how to solve a large class of robust nonlinear optimization problems. The fascinating and appealing property of our approach is that any convex uncertainty set can be used. We give an explicit formulation of the dual of a robust nonlinear optimization problem, which contains the convex conjugate functions of the objective and constraint functions of the (deterministic) primal, and the perspectives of the convex functions that define the uncertainty set. Given an optimal solution of this dual problem and a corresponding KKT vector, we show how to recover the primal optimal solution. We obtain computationally tractable robust counterparts for many new robust nonlinear optimization problems, including problems with robust quadratic constraints, second order cone constraints, and SOS-convex polynomials.

keywords: robust optimization; general convex uncertainty regions; linear optimization
MSC classification: 90C30

1 Introduction

Optimization problems are often affected by uncertainty. A slight change in the parameters of the problem may render a previously optimal solution infeasible or suboptimal (Ben-Tal et al. 2009, p. ix). Robust Optimization (RO) is a method that avoids infeasibility and decay of the solution (Ben-Tal et al. 2009, Bertsimas et al. 2011). In basic versions of RO, the constraints have to hold for all parameter realizations in a prespecified (infinite) uncertainty set. Let us focus on the following Robust Counterpart (RC), where $\boldsymbol{a^i} \in \mathbb{R}^L$ is an uncertain parameter and

[1]Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School. CentER Junior Extramural Fellow, Tilburg University. E-mail: `b.l.gorissen@tilburguniversity.edu`
[2]Department of Econometrics and Operations Research, Tilburg University.

$f_i : \mathbb{R}^L \times \mathbb{R}^n \to \mathbb{R}$ are the constraint functions:

$$
\begin{aligned}
\text{(RC)} \quad & \sup_{\boldsymbol{x}\in\mathbb{R}^n} \quad \boldsymbol{c}^\top \boldsymbol{x} \\
& \text{s.t.} \quad f_i(\boldsymbol{a^i}, \boldsymbol{x}) \leq 0 \quad \forall \boldsymbol{a^i} \in \mathcal{U}_i \quad \forall i \in I. \qquad (1)
\end{aligned}
$$

This formulation is called "robust" since the constraints have to hold for all $\boldsymbol{a^i}$ in the uncertainty set $\mathcal{U}_i \subset \mathbb{R}^L$. Note that we assume the objective to be certain, which is without any loss of generality, since an uncertain objective can be turned into a certain objective and an uncertain constraint via an epigraph reformulation. The assumption that the uncertainty set is constraint-wise, is also made without any loss of generality (Ben-Tal et al. 2009, p. 11). The formulation (1) is not tractable in its current form, since it has infinitely many constraints. Part of the research in RO is dedicated to finding an equivalent formulation with finitely many variables and constraints. The current general results reformulate each constraint based on duality in $\boldsymbol{a^i}$ (Ben-Tal et al. 2009, Bertsimas et al. 2011). For example, a linear constraint with polyhedral uncertainty can be reformulated by applying LP duality:

$$
(\boldsymbol{a^i})^\top \boldsymbol{x} \leq b_i \; \forall \boldsymbol{a^i} : \boldsymbol{D}\boldsymbol{a^i} \leq \boldsymbol{d} \;\Leftrightarrow\; \sup_{\boldsymbol{a^i}:\boldsymbol{D}\boldsymbol{a^i}\leq\boldsymbol{d}} (\boldsymbol{a^i})^\top \boldsymbol{x} \leq b_i \;\Leftrightarrow\; \inf_{\boldsymbol{y^i}\geq\boldsymbol{0}:\boldsymbol{D}^\top\boldsymbol{y^i}=\boldsymbol{x}} \boldsymbol{d}^\top \boldsymbol{y^i} \leq b_i. \qquad (2)
$$

The inf operator can now be omitted, since if the inequality holds for one $\boldsymbol{y^i}$, it holds for the infimum. In this work, we take a different approach, extending our previous results on robust *linear* optimization (Gorissen et al. 2014). Instead of dualizing each constraint w.r.t. $\boldsymbol{a_i}$, we derive an explicit formulation of the dual of (RC) w.r.t. $\boldsymbol{x}$ (Section 2).

The innovativeness with respect to Ben-Tal et al. (2015) is twofold. First, solving the dual offers many advantages, such as the ability to deal with any convex uncertainty set, and the dual may be more efficient to solve or easier to formulate. For more details on these advantages see Section 3. Second, we are the first to derive generic results for robust quadratic, second-order cone and SOS-convex constraints; for more details see Sections 4.2–4.4. Especially the results for robust quadratic and second-order constraints are a significant extension of the existing literature.

Our method is demonstrated on an illustrative example in Section 4.6.

2 The dual of (RC)

We assume that I is finite, and for each i, that $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ is concave in $\boldsymbol{a^i}$ on $\mathcal{U}_i$ (for each fixed $\boldsymbol{x}$ in $\mathbb{R}^n$) and closed proper convex in $\boldsymbol{x}$ (for each fixed $\boldsymbol{a^i}$ in $\mathcal{U}_i$), and that $\mathcal{U}_i = \{\boldsymbol{a^i} \in \mathbb{R}^L : g_{ik}(\boldsymbol{a^i}) \leq 0 \quad \forall k \in K_i\}$ is a bounded uncertainty region, where $g_{ik} : \mathbb{R}^L \to \mathbb{R}$ is convex for each i and k, and K_i is finite.

We define the concave conjugate of $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ w.r.t. the first argument, the convex conjugate of $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ w.r.t. the second argument, and the convex conjugate of $g_{ik}(\boldsymbol{a^i})$ by, respectively:

$$(f_i)_*(\boldsymbol{v^i}, \boldsymbol{x}) = \inf_{\boldsymbol{a^i} \in \mathbb{R}^L} \{(\boldsymbol{v^i})^\top \boldsymbol{a^i} - f_i(\boldsymbol{a^i}, \boldsymbol{x})\}, \tag{3}$$

$$f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{(\boldsymbol{u^i})^\top \boldsymbol{x} - f_i(\boldsymbol{a^i}, \boldsymbol{x})\}, \text{ and} \tag{4}$$

$$g_{ik}^*(\boldsymbol{u^i}) = \sup_{\boldsymbol{a^i} \in \mathbb{R}^L} \{(\boldsymbol{u^i})^\top \boldsymbol{a^i} - g_{ik}(\boldsymbol{a^i})\}.$$

The concave conjugate of f is jointly concave in $(\boldsymbol{v^i}, \boldsymbol{x})$, since it is the infimum of functions that are jointly concave in $(\boldsymbol{v^i}, \boldsymbol{x})$. Similarly, the convex conjugates of f_i and g_{ik} are jointly convex. Convex (or concave) conjugates provide a description of the convex (or concave) envelope of a function in terms of the supporting hyperplances. For an introduction to conjugate functions and identities for deriving conjugates, we refer to Boyd and Vandenberghe (2004, §3.3), Rockafellar (1970, §16) and Ben-Tal et al. (2015). The perspective of the function g_{ik} is defined by $h_{ik} : \mathbb{R}^L \times \mathbb{R}_+$, $h_{ik}(\boldsymbol{a^i}, y_i) = y_i g_{ik}(\boldsymbol{a^i}/y_i)$ for $y_i > 0$ and $h_{ik}(\boldsymbol{a^i}, 0) = 0$ if $\boldsymbol{a^i} = 0$, ∞ otherwise. For brevity and clarity, we will not introduce specific functions for each perspective. Instead, we denote the perspective using division by y_i, with the convention that for $y_i = 0$, the function value is either 0 or ∞ depending on whether the arguments divided by y_i are all zero:

$$y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i} \right) = \begin{cases} y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i} \right) & \text{if } y_i > 0 \\ 0 & \text{if } y_i = 0,\ \boldsymbol{v^i} = \boldsymbol{0} \text{ and } \boldsymbol{w^i} = \boldsymbol{0} \\ \infty & \text{else.} \end{cases} \tag{5}$$

The perspective of a convex function is convex (Rockafellar 1970, p. 35).

Let $\boldsymbol{a^i}$ be an element of $\mathcal{U}_i$, and assume $\cap_i$ ri(dom$(f_i(\boldsymbol{a^i}, \cdot))$) is not empty, where ri denotes the relative interior and dom$(f_i(\boldsymbol{a^i}, \cdot)) = \{\boldsymbol{x} : f_i(\boldsymbol{a^i}, \boldsymbol{x}) < \infty\}$. The dual of $\sup_{\boldsymbol{x} \in \mathbb{R}^n} \{\boldsymbol{c}^\top \boldsymbol{x} :$

$f_i(\boldsymbol{a^i}, \boldsymbol{x}) \leq 0 \quad \forall i \in I\}$ is given by (cf. Boyd and Vandenberghe 2004, §5.7.1):

$$\inf_{\boldsymbol{y} \geq \boldsymbol{0}} \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \boldsymbol{c}^\top \boldsymbol{x} - \sum_{i \in I} y_i f_i \left(\boldsymbol{a^i}, \boldsymbol{x} \right) \right\} = \inf_{\boldsymbol{y} \geq \boldsymbol{0}} \left(\sum_{i \in I} y_i f_i \right)^* \left(\boldsymbol{a^i}, \boldsymbol{c} \right)$$
$$= \inf_{\boldsymbol{y} \geq \boldsymbol{0}, \boldsymbol{w^i} \in \mathbb{R}^n} \left\{ \sum_{i \in I} y_i f_i^* \left(\boldsymbol{a^i}, \frac{\boldsymbol{w^i}}{y_i} \right) : \sum_{i \in I} \boldsymbol{w^i} = \boldsymbol{c} \right\},$$

where the first equality uses the definition of the conjugate, and the second equality is based on the sum rule for conjugates and the rule for the conjugate of a scalar multiplied with a function (Rockafellar 1970, Thm. 16.1 and 16.4). The optimistic dual is obtained by additionally optimizing over $\boldsymbol{a^i}$ in the uncertainty set:

$$\text{(OD)} \quad \inf_{\boldsymbol{y} \geq \boldsymbol{0},\, \boldsymbol{a^i} \in \mathbb{R}^L,\, \boldsymbol{w^i} \in \mathbb{R}^n} \sum_{i \in I} y_i f_i^* \left(\boldsymbol{a^i}, \frac{\boldsymbol{w^i}}{y_i} \right)$$
$$\text{s.t.} \quad \sum_{i \in I} \boldsymbol{w^i} = \boldsymbol{c}$$
$$g_{ik} \left(\boldsymbol{a^i} \right) \leq 0 \quad \forall i \in I \quad \forall k \in K_i. \tag{6}$$

Beck and Ben-Tal (2009) have shown that (OD) is the dual of (RC), without giving an explicit formulation of (OD), and that strong duality holds if (RC) satisfies the Slater condition.

Unfortunately, (OD) is difficult to solve due to the nonconvexity introduced by the product $y_i f_i^*(\boldsymbol{a^i}, \boldsymbol{w^i}/y_i)$. The equivalent convex problem (COD) is obtained by substituting $y_i \boldsymbol{a^i} = \boldsymbol{v^i}$ and multiplying constraint (6) with y_i (cf. Gorissen et al. 2014):

$$\text{(COD)} \quad \inf_{\boldsymbol{y} \geq \boldsymbol{0}, \boldsymbol{v^i} \in \mathbb{R}^L, \boldsymbol{w^i} \in \mathbb{R}^n} \sum_{i \in I} y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i} \right)$$
$$\text{s.t.} \quad \sum_{i \in I} \boldsymbol{w^i} = \boldsymbol{c} \tag{7}$$
$$y_i g_{ik} \left(\frac{\boldsymbol{v^i}}{y_i} \right) \leq 0 \quad \forall i \in I \quad \forall k \in K_i. \tag{8}$$

Note that (COD) is a convex optimization problem. The following theorem shows how (COD) provides an optimal solution of (RC). It is an extension of the theorem in Gorissen et al. (2014) for robust linear optimization.

Theorem 1 *Assume (COD) satisfies the Slater condition, then (COD) and (RC) have the same optimal value. Moreover, the part of the KKT vector of (COD) that corresponds to constraint*

(7) *gives an optimal solution of (RC).*

Proof. Since $f_i^{**} = f_i$, because f_i is closed and convex, constraint (1) can be written as:

$$\sup_{\boldsymbol{u^i} \in \mathbb{R}^n} \{(\boldsymbol{u^i})^\top \boldsymbol{x} - f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i})\} \leq 0 \quad \forall \boldsymbol{a^i} \in \mathcal{U}_i \quad \forall i \in I,$$

or, equivalently:

$$(\boldsymbol{u^i})^\top \boldsymbol{x} \leq f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i}) \quad \forall \boldsymbol{a^i} \in \mathcal{U}_i \quad \forall \boldsymbol{u^i} \in \mathbb{R}^n \quad \forall i \in I. \tag{9}$$

This reformulation of constraint (1) is well known outside the field of RO, see, e.g., Mehrotra and Papp (2014). Constraint (9) holds if and only if:

$$(\boldsymbol{u^i})^\top \boldsymbol{x} \leq b_i \quad \forall \boldsymbol{a^i} \in \mathcal{U}_i \quad \forall \boldsymbol{u^i} \in \mathbb{R}^n \quad \forall b_i \in \mathbb{R} : b_i \geq f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i}) \quad \forall i \in I. \tag{10}$$

By substituting constraint (10) into (RC), (RC) becomes a linear optimization problem with a convex uncertainty set. The optimistic dual can be obtained using Gorissen et al. (2014), and is exactly the same as (OD), and therefore equivalent to the convex optimization problem (COD).

By Theorem 1 of Gorissen et al. (2014), an optimal solution of (RC) is given by a KKT vector of constraint (7) only if the uncertainty set is bounded. Unfortunately, in our case $\boldsymbol{u^i}$ is unbounded. The result by Gorissen et al. still holds if the *substitution equivalence condition* holds, i.e., if $y_i = 0$ implies both $\boldsymbol{v^i} = \boldsymbol{0}$ and $\boldsymbol{w^i} = \boldsymbol{0}$ for an optimal solution of (COD). This implication follows from the definition of $0 f_i^*(\boldsymbol{v^i}/0, \boldsymbol{w^i}/0)$ in (5). ■

Seemingly, the initial assumption that $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ is concave in $\boldsymbol{a^i}$ (for each fixed $\boldsymbol{x}$) is overly restrictive, since it was only used to ensure that f_i^* is jointly convex. One may wonder if $f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i})$ can be jointly convex while $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ is not concave in $\boldsymbol{a^i}$. The answer turns out to be negative. Assume that f_i^* is jointly convex. Then, for fixed $\boldsymbol{x}$, the function $h_{\boldsymbol{x}}(\boldsymbol{a^i}, \boldsymbol{u^i}) = (\boldsymbol{u^i})^\top \boldsymbol{x} - f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i})$ is jointly concave. Since $f_i(\boldsymbol{a^i}, \boldsymbol{x}) = f_i^{**}(\boldsymbol{a^i}, \boldsymbol{x}) = \sup_{\boldsymbol{u^i} \in \mathbb{R}^n} h_{\boldsymbol{x}}(\boldsymbol{a^i}, \boldsymbol{u^i})$, it follows that $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ is concave in $\boldsymbol{a^i}$ (Boyd and Vandenberghe 2004, §3.2.5).

When the functions g_{ik} that define the uncertainty set are explicitly given, it is straightforward to formulate their perspectives. The following lemma provides the formulation of the perspective when the uncertainty set is conic representable.

Lemma 1 *Let $\mathcal{K}_i$ be a closed convex cone, and suppose the uncertainty region $\mathcal{U}_i$ is described by one or more conic inclusion constraints: $\boldsymbol{D^i a^i} - \boldsymbol{d^i} \in \mathcal{K}_i$ for a given matrix $\boldsymbol{D^i} \in \mathbb{R}^{m\times L}$ and a given vector $\boldsymbol{d^i} \in \mathbb{R}^m$. Then, the corresponding perspective in constraint* (8) *is $\boldsymbol{D^i v^i} - y_i \boldsymbol{d^i} \in \mathcal{K}_i$.*

Proof. Define $g_{ik}(\boldsymbol{a^i})$ as an indicator function, taking the value 0 if $\boldsymbol{a^i}$ satisfies the conic inclusion constraint and ∞ otherwise. When $y_i > 0$, $y_i g_{ik}(\boldsymbol{v^i}/y_i) \le 0$ if and only if $g_{ik}(\boldsymbol{v^i}/y_i) = 0$, i.e., $\boldsymbol{D^i v^i}/y_i - \boldsymbol{d^i} \in \mathcal{K}_i$. Since $\mathcal{K}_i$ is a cone, this is equivalent to $\boldsymbol{D^i v^i} - y_i \boldsymbol{d^i} \in \mathcal{K}_i$. When $y_i = 0$, $y_i g_{ik}(\boldsymbol{v^i}/y_i) \le 0$ if and only if there exists an $y_i^* > 0$ such that $g_{ik}(\boldsymbol{v^i}/\varepsilon) = 0$ for all ε in the interval $(0, y_i^*]$. So, for all ε in $(0, y_i^*]$, $\boldsymbol{D^i v^i} - \varepsilon \boldsymbol{d^i} \in \mathcal{K}_i$. Since $\mathcal{K}_i$ is closed, $\boldsymbol{D^i v^i} - y_i \boldsymbol{d^i} \in \mathcal{K}_i$. ■

3 Advantages of our method

We distinguish between advantages of formulating and of solving the dual problem.

3.1 Advantages of formulating the dual problem

The two advantages of formulating the dual problem is that it can be done for any convex uncertainty set, and that is often easier than formulating the primal problem. To demonstrate this, recall that the prevalent (primal) approach dualizes each constraint by considering the left hand side of each constraint (1) as a maximization problem over $\boldsymbol{a^i}$ as shown in (2). Assuming $\text{ri}(\text{dom}(\mathcal{U}_i)) \cap \text{ri}(\text{dom}(f_i(\cdot, \boldsymbol{x}))) \ne \emptyset$, dualizing each constraint of (RC) using Fenchel's duality theorem yields the following equivalent convex optimization problem (Ben-Tal et al. 2015):

$$\begin{aligned} \sup_{\boldsymbol{x}\in\mathbb{R}^n,\, \boldsymbol{v^{ik}}\in\mathbb{R}^L,\, \boldsymbol{\lambda}\ge 0} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \sum_{k\in K_i} \lambda_{ik} g_{ik}^* \left(\frac{\boldsymbol{v^{ik}}}{\lambda_{ik}}\right) - (f_i)_* \left(\sum_{k\in K_i} \boldsymbol{v^{ik}}, \boldsymbol{x}\right) \le 0 \quad \forall i \in I. \end{aligned}$$

For this result to be usable, the convex conjugate of $g_{ik}(\boldsymbol{a^i})$ needs to have a closed-form expression, or it needs to be the optimal value of a convex minimization problem with finitely many objectives and constraints. Many examples of the latter are given by Ben-Tal et al. (2015). A similar condition is necessary for the concave conjugate of $f_i(\boldsymbol{a^i}, \boldsymbol{x})$. From now on, we say that "a tractable conjugate exists" to indicate that these conditions are satisfied.

Not every function has a tractable conjugate, e.g., $f(x) = \exp(x^2)$ if $x \ge \sqrt{2}/2$, $f(x) = \infty$ otherwise. Many (convex) uncertainty regions therefore previously led to intractable RCs for

Table 1: Eight different cases for the existence of tractable conjugates. The method by Ben-Tal et al. (2015) can solve cases 7 and 8, while our method can solve cases 2, 5, 6 and 8

Case	Tractable $f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i})$ exists	Tractable $(f_i)_*(\boldsymbol{u^i}, \boldsymbol{x})$ exists	Tractable $g_{ik}^*(\boldsymbol{u^i})$ exists
1	-	-	-
2	✓	-	-
3	-	✓	-
4	-	-	✓
5	✓	✓	-
6	✓	-	✓
7	-	✓	✓
8	✓	✓	✓

uncertain nonlinear optimization problems, but can now be solved with our method. Examples of these are given in Gorissen et al. (2014), such as uncertainty sets for probability vectors (e.g., based on ϕ-divergence, Rényi divergence or Bregman distance) or uncertainty sets that are uncertain themselves.

Table 1 shows the applicability of our method on the eight different cases for the existence of tractable conjugates. The method by Ben-Tal et al. (2015) can solve cases 7 and 8, while our method can solve cases 2, 5, 6 and 8. All eight cases can be solved with a cutting plane method (e.g., Bertsimas et al. 2014), but their performance is not yet well understood. Accordingly, to the best of our knowledge, the method described in this paper is the only method that can give exact reformulations for the following two classes of problems: (1) problems for which a tractable convex conjugate of $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ exists, but a tractable concave conjugate of $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ does not necessarily exist; and (2) problems for which the uncertainty set is any convex set represented by convex functions, even if tractable conjugates of $g_{ik}(\boldsymbol{a^i})$ do not exist.

For the cases that can be solved with our method, the primal requires the support function of the uncertainty set, which is harder to determine than the perspectives required by our method. there is more flexibility in the uncertainty regions that can be considered.

3.2 Advantages of solving the dual problem

The dual may be more efficient to solve than the primal. This can be due to (1) a different number of variables and constraints, (2) a lower self-concordance parameter, (3) better sparsity structure, (4) numerical instability of the primal when variables approach their bounds, (5) linear constraints in the dual (such as in geometric optimization).

When both the primal and the dual problem cannot be solved to optimality, the duality gap gives an optimality guarantee.

Solving the dual problem has two disadvantages (Gorissen et al. 2014). First, problems with integer variables cannot be dualized. However, this is not a serious issue, since our method can solve continuous relaxations and can therefore be used in branch & bound solvers. Second, our method requires a KKT vector to recover the primal solution. Interior point methods and SQP solvers compute these vectors as part of their algorithm. For other solvers, a KKT vector can easily be derived for an optimal solution.

4 Examples of conjugates

Our method requires tractable convex conjugates of the constraint functions. In this section, we derive these conjugates for many optimization problems, such as uncertain quadratic constraints, SOS-convex polynomial constraints, and second order conic constraints, giving proof of the broad applicability of our method. The expressions for many conjugate functions that we derive here have not been published yet, so this section also serves as a reference, for example when using the method by Ben-Tal et al. (2015). Surprisingly, the conjugates are often conic representable even if the uncertain parameters get added as optimization variables.

Since we focus on conjugate functions and not on specific constraints, let us drop the subscript or superscript i for convenience. Although our method requires the perspective of the conjugate, formulating the perspective is straightforward for each of the examples.

We stress that for formulating (COD), not all constraints have to be of the same type. So, (RC) may have a mixture of any of the following constraints.

4.1 Linear in the optimization variables

Vector $\boldsymbol{x}$. Let $\boldsymbol{h} : \mathbb{R}^L \to \mathbb{R}^n$ be a concave function, and let $f(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{a})^\top \boldsymbol{x}$ if $\boldsymbol{x} \geq \mathbf{0}$, $f(\boldsymbol{a}, \boldsymbol{x}) = \infty$ otherwise. The convex conjugate of $f(\boldsymbol{a}, \boldsymbol{x})$ is:

$$f^*(\boldsymbol{a}, \boldsymbol{u}) = \begin{cases} 0 & \text{if } \boldsymbol{u} \leq \boldsymbol{h}(\boldsymbol{a}) \\ \infty & \text{otherwise.} \end{cases}$$

Hence, in (COD) the corresponding term in the objective is zero (if $y > 0$ or both $\boldsymbol{v} = \mathbf{0}$ and

$\boldsymbol{w} = \mathbf{0}$), and (COD) gets the additional constraint $\boldsymbol{w} \leq y\boldsymbol{h}(\boldsymbol{v}/y)$.

If the concave conjugate of $h(\boldsymbol{a})$ does not have a tractable expression, the method by Ben-Tal et al. (2015) fails to produce a tractable reformulation of RC, whereas our method can be applied for any concave function h. This result can also be obtained from Gorissen et al. (2014) by rewriting the constraint as $\boldsymbol{u}^\top \boldsymbol{x} \leq 0 \ \forall \boldsymbol{a} \in \mathcal{U} \ \forall \boldsymbol{u} \in \mathbb{R}^n : (\boldsymbol{u})^{LB} \leq \boldsymbol{u} \leq \boldsymbol{h}(\boldsymbol{a})$, where $(\boldsymbol{u})^{LB}$ is a suitably chosen lower bound that makes the uncertainty set bounded without modifying the original constraint (e.g., take the j^{th} component equal to $\min_{\boldsymbol{a} \in \mathcal{U}} h_j(\boldsymbol{a})$).

Positive semidefinite $\boldsymbol{X}$. Suppose $f(\boldsymbol{a}, b, \boldsymbol{X}) = b - \boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{a}$ if $\boldsymbol{X} \in \mathbb{S}^n_+$, ∞ otherwise ($\mathbb{S}^n_+$ being the cone of positive semidefinite $n \times n$ matrices). The convex conjugate of $f(\boldsymbol{a}, b, \boldsymbol{X})$ is given by:

$$f^*(\boldsymbol{a}, b, \boldsymbol{U}) = \begin{cases} -b & \text{if } -\boldsymbol{U} - \boldsymbol{a}\boldsymbol{a}^\top \in \mathbb{S}^n_+ \\ \infty & \text{otherwise.} \end{cases}$$

A semidefinite representation (SDr) of the constraint $-\boldsymbol{U} - \boldsymbol{a}\boldsymbol{a}^\top \in \mathbb{S}^n_+$ can be obtained via the Schur complement:

$$\begin{pmatrix} 1 & \boldsymbol{a}^\top \\ \boldsymbol{a} & -\boldsymbol{U} \end{pmatrix} \in \mathbb{S}^{n+1}_+.$$

For this problem, a tractable reformulation can also be derived by taking the concave conjugate of $f(\boldsymbol{a}, b, \boldsymbol{X})$ w.r.t. $\boldsymbol{a}$ and b. A tractable expression for this conjugate does not exist in the literature, but the derivation is similar to the one in Section 4.2 and gives an SDr. The advantage of our approach is that it can deal with general convex uncertainty sets.

When the uncertainty set is ellipsoidal and $\boldsymbol{X}$ is not restricted to be positive semidefinite, the RC is known to be SDr (Ben-Tal et al. 2009, p. 24). The first step in the derivation of the RC is to change the uncertain parameter from $\boldsymbol{a}$ to $(\boldsymbol{a}, \boldsymbol{a}\boldsymbol{a}^\top)$, so the constraint becomes linear in the uncertain parameters (and is still linear in $\boldsymbol{X}$). Then, the uncertainty set is replaced with its convex hull, which is SDr. We would also have to take these steps to obtain the dual.

A constraint that is quadratic in the uncertain parameters may appear when creating a meta-model of a black-box function $h : \mathbb{R}^n \to \mathbb{R}$. Suppose m pairs of observations $(\boldsymbol{\nu^i}, h(\boldsymbol{\nu^i})) = (\boldsymbol{\nu^i}, y_i)$

$(i = 1, \ldots, m)$ are given, and h can be approximated with a quadratic function:

$$h(\boldsymbol{\nu^i}) \approx (\boldsymbol{\nu^i})^\top \boldsymbol{\Gamma} \boldsymbol{\nu^i} + \boldsymbol{\beta}^\top \boldsymbol{\nu^i} + \gamma.$$

A reasonable way of estimating the parameters $\boldsymbol{\Gamma}$, $\boldsymbol{\beta}$ and γ is to minimize the estimation error. There are many convex functions that can be used as distance measures, and the particular choice is not so relevant for this example, so let us consider the least squares estimates:

$$\min_{\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}, \boldsymbol{\beta} \in \mathbb{R}^n, \gamma \in \mathbb{R}} \quad \sum_{i=1}^{m} \left(y_i - (\boldsymbol{\nu^i})^\top \boldsymbol{\Gamma} \boldsymbol{\nu^i} - \boldsymbol{\beta}^\top \boldsymbol{\nu^i} - \gamma \right)^2.$$

Here, $\boldsymbol{\nu^i}$ are parameters. It may be known that the function h is convex and nonnegative on a set $S \subset \mathbb{R}^n$, and it is desired that the estimate also satisfies these properties (Siem et al. 2008). This can be enforced by $\boldsymbol{\Gamma} \in \mathbb{S}^n_+$ and:

$$\boldsymbol{\nu}^\top \boldsymbol{\Gamma} \boldsymbol{\nu} + \boldsymbol{\beta}^\top \boldsymbol{\nu} + \gamma \geq 0 \quad \forall \boldsymbol{\nu} \in S. \tag{11}$$

Previously, this constraint could be reformulated via the S-lemma when S is ellipsoidal, or approximated using sums of squares if S is a semi-algebraic set (Siem et al. 2008). We can solve problems with the nonnegativity constraint (11) for any convex set S defined by finitely many convex functions.

4.2 Quadratic in the optimization variables

Consider the following two constraints:

$$\boldsymbol{x}^\top \boldsymbol{P}(\boldsymbol{\zeta})^\top \boldsymbol{P}(\boldsymbol{\zeta}) \boldsymbol{x} + \boldsymbol{b}(\boldsymbol{\zeta})^\top \boldsymbol{x} + c(\boldsymbol{\zeta}) \leq 0 \qquad \forall \boldsymbol{\zeta} \in \mathcal{Z} \tag{12}$$

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \leq 0 \qquad \forall (\boldsymbol{A}, \boldsymbol{b}, c) \in \mathcal{U}. \tag{13}$$

Ben-Tal et al. (2009, Chapter 6) focus on constraint (12) where the parameters are assumed to be affine in $\boldsymbol{\zeta}$. This constraint is not concave in the uncertain parameter, and can only be reformulated to a small number of convex constraints in special cases, e.g., when $\mathcal{Z}$ is the convex hull of a small number of scenarios or when $\mathcal{Z}$ is the set of all matrices whose maximal singular value is 1.

We therefore consider constraint (13), which offers the possibility to formulate the uncertainty on $\boldsymbol{A}$ directly, which may be more natural in some cases than describing the uncertainty via $\boldsymbol{P}$. We are the first to show that (13) can be solved for any convex uncertainty set. Our result generalizes the specific results by Goldfarb and Iyengar (2003), who found a CQr RC for two cases:

1. when $(\boldsymbol{A}, \boldsymbol{b}, c) = \sum_{j=1}^{k} s_j(\boldsymbol{A^j}, \boldsymbol{b^j}, c_j)$ where A^j, $\boldsymbol{b^j}$ and c_j are fixed, $\boldsymbol{A^j}$ is positive semidefinite and $\boldsymbol{s}$ is nonnegative and either norm bounded or in a polyhedron, and
2. when $\boldsymbol{A} = \boldsymbol{V}^\top \boldsymbol{F} \boldsymbol{V}$ where both matrices $\boldsymbol{V}$ and $\boldsymbol{F}$ are uncertain. The uncertainty set for $\boldsymbol{V}$ binds the norm of the columns, whereas the uncertainty set on $\boldsymbol{F}$ is defined in terms of $\boldsymbol{F}^{-1}$.

For the first case we also get a CQr, but we do not require an affine parameterization for $\boldsymbol{b^j}$ and c_j or a specific uncertainty set for $\boldsymbol{s}$. For the second case, our method can be used by making $\boldsymbol{A} \preceq \boldsymbol{V}^\top \boldsymbol{F} \boldsymbol{V}$ part of the uncertainty set, which is SDr via the Schur complement, and with any convex constraint on $\boldsymbol{V}$ or $\boldsymbol{F}^{-1}$. In contrast to Goldfarb and Iyengar (2003), we do not impose a specific structure on $\boldsymbol{A}$, other than that it is contained in a convex set.

Let $\mathcal{U} \subset \mathbb{S}^n_+ \times \mathbb{R}^n \times \mathbb{R}$, and let $f(\boldsymbol{A}, \boldsymbol{b}, c, \boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c$. We show that the convex conjugate is SDr, and sometimes conic quadratic representable (CQr). The derivation of the conjugate and an SDr follow from Boyd and Vandenberghe (2004, Appendix A.5.5):

$$f^*(\boldsymbol{A}, \boldsymbol{b}, c, \boldsymbol{u}) = \begin{cases} \frac{1}{4}(\boldsymbol{u} - \boldsymbol{b})^\top \boldsymbol{A}^\dagger (\boldsymbol{u} - \boldsymbol{b}) - c, & \text{if } \exists \boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{u} - \boldsymbol{b} = \boldsymbol{A}\boldsymbol{y} \\ \infty & \text{otherwise,} \end{cases}$$

where $\boldsymbol{A}^\dagger = \boldsymbol{A}^{-1}$ if $\boldsymbol{A}$ is invertible, and a generalized inverse otherwise. An SDr can be obtained via the Schur complement: $y f^*(\boldsymbol{A}/y, \boldsymbol{b}/y, c/y, \boldsymbol{u}/y) \leq t$ is equivalent to:

$$\begin{pmatrix} 4\boldsymbol{A} & \boldsymbol{u} - \boldsymbol{b} \\ (\boldsymbol{u} - \boldsymbol{b})^\top & c + t \end{pmatrix} \in \mathbb{S}^{n+1}_+.$$

A more tractable result can be obtained when $\boldsymbol{A}$ is affinely parameterized by an uncertain nonnegative vector $\boldsymbol{s} \in S \subset \mathbb{R}^k_+$:

$$\boldsymbol{A} = \sum_{j=1}^{k} s_j \boldsymbol{A^j}, \tag{14}$$

where the (fixed) matrices $\boldsymbol{A^j}$ are positive semidefinite with rank k_j: $\boldsymbol{A^j} = (\boldsymbol{D^j})^\top \boldsymbol{D^j}$ for a $k_j \times n$ matrix $\boldsymbol{D^j}$. A CQr of $yf^*(\boldsymbol{s}/y, \boldsymbol{b}/y, c/y, \boldsymbol{u}/y) \leq t$ is given by (Nesterov and Nemirovskii 1994, §6.3.1):

$$\begin{cases} \left\| \begin{pmatrix} 2\pi_j \\ \tau_j - s_j \end{pmatrix} \right\|_2 \leq \tau_j + s_j \quad j = 1, \ldots, k \\ \sum_{j=1}^k (\boldsymbol{D^j})^\top \pi_j = \frac{1}{2}(\boldsymbol{u} - \boldsymbol{b}) \\ \sum_{j=1}^k \tau_j \leq c + t \\ \pi \in \mathbb{R}^k, \ \tau \in \mathbb{R}^k. \end{cases}$$

Example (variance). Suppose $\boldsymbol{a}$ is a probability vector on $\boldsymbol{x}$, i.e., $\mathcal{U} \subseteq \Delta^{n-1}$ (the standard simplex in $\mathbb{R}^n$). Postek et al. (2014) showed that the variance of $\boldsymbol{x}$ can be written as $\min_{\pi \in \mathbb{R}} \sum_{j=1}^n a_j (x_j - \pi)^2$ (Postek et al. 2014), and that when $\mathcal{U}$ is bounded:

$$\max_{\boldsymbol{a} \in \mathcal{U}} \min_{\pi \in \mathbb{R}} \sum_{j=1}^n a_j (x_j - \pi)^2 = \min_{\pi \in \mathbb{R}} \max_{\boldsymbol{a} \in \mathcal{U}} \sum_{j=1}^n a_j (x_j - \pi)^2.$$

Therefore, the constraint $f(\boldsymbol{a}, \boldsymbol{x}, \pi) = \sum_{j=1}^n a_j (x_j - \pi)^2 - d \leq 0$, where π is an optimization variable, is equivalent to binding the variance from above by d. This constraint can be expressed as the following quadratic form with an arrow matrix:

$$f(\boldsymbol{a}, \boldsymbol{x}, y) = \begin{pmatrix} \boldsymbol{x} \\ \pi \end{pmatrix}^\top \begin{pmatrix} a_1 & & & -a_1 \\ & \ddots & & \vdots \\ & & a_n & -a_n \\ -a_1 & \ldots & -a_n & \sum_{j=1}^n a_j \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \pi \end{pmatrix} - d.$$

The arrow matrix can be written as $\sum_{j=1}^n a_j (\boldsymbol{D^j})^\top \boldsymbol{D^j}$ for $1 \times n$ vectors $\boldsymbol{D^j}$, that have a 1 at position j, -1 at position n and zeros elsewhere. The convex conjugate (w.r.t. both $\boldsymbol{x}$ and π) is therefore CQr.

4.3 Linear in the uncertain parameters

Let $\mathcal{U} \subset \mathbb{R}^L_+$, let $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^L$ be a convex function, and let $f(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{h}(\boldsymbol{x}) = \sum_{j=1}^L a_j h_j(\boldsymbol{x})$. The conjugate of $f(\boldsymbol{a}, \boldsymbol{x})$ is:

$$f^*(\boldsymbol{a}, \boldsymbol{u}) = \inf_{\boldsymbol{w^j} \in \mathbb{R}^n} \left\{ \sum_{j=1}^L a_j h_j^* \left(\frac{\boldsymbol{w^j}}{a_j} \right) : \sum_{j=1}^L \boldsymbol{w^j} = \boldsymbol{u} \right\}.$$

We now specialize this result to a specific choice for h.

SOS-convex polynomial optimization. Let us first recall some definitions about *sum of squares* (SOS). For a more detailed description, see Laurent (2009). A polynomial $p(\boldsymbol{x}) = \sum_{i=1}^n \prod_{j=1}^m x_{ij}^{c_{ij}}$ with $c_{ij} \in \mathbb{N} \cup \{0\}$ has degree d if $\sum_{j=1}^m c_{ij} \leq d$ $(i = 1, \ldots, n)$, and is SOS if $p(\boldsymbol{x}) = \boldsymbol{q}(\boldsymbol{x})^\top \boldsymbol{q}(\boldsymbol{x})$ for some multivariate polynomial $\boldsymbol{q}$. For example, $2x_1^2 - 4x_1x_2 + 4x_2^2$ is SOS since it equals $x_1^2 + (x_1 - 2x_2)^2$. A multivariate polynomial is SOS-convex if its Hessian equals $\boldsymbol{Q}(\boldsymbol{x})\boldsymbol{Q}(\boldsymbol{x})^\top$ for some matrix $\boldsymbol{Q}(\boldsymbol{x})$ whose entries are polynomial in $\boldsymbol{x}$. An SOS-convex polynomial is convex, but a convex polynomial is not necessarily SOS-convex (Ahmadi and Parrilo 2013). The class of SOS polynomials $p(\boldsymbol{x})$ of degree at most d is denoted by Σ_d^2, which has a semidefinite representation. Similary, SOS-convexity has a semidefinite representation.

Suppose $h_j(\boldsymbol{x})$ is an SOS-convex polynomial of degree d. Then,

$$h_j^*(\boldsymbol{u^j}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{(\boldsymbol{u^j})^\top \boldsymbol{x} - h_j(\boldsymbol{x})\} \leq t$$

if and only if:

$$p_t(\boldsymbol{x}) = t + h_j(\boldsymbol{x}) - (\boldsymbol{u^j})^\top \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \in \mathbb{R}^n. \tag{15}$$

Constraint (15) clearly holds if $p_t(\boldsymbol{x})$ is SOS. We now show that this is also a necessary condition. Note that $p_t(\boldsymbol{x})$ is SOS-convex since its Hessian equals that of $h_j(\boldsymbol{x})$. Let $t = t^*$ be the smallest t for which inequality (15) holds. Then there exists some $\boldsymbol{x}^*$ in $\mathbb{R}^n$ for which $p_{t^*}(\boldsymbol{x}^*) = 0$. Since $\boldsymbol{x}^*$ is a minimizer, $\nabla p_{t^*}(\boldsymbol{x}^*) = 0$. By Lemma 8 of Helton and Nie (2010), constraint (15) holds if and only if $p_{t^*}(\boldsymbol{x})$ is SOS. Clearly, $p_t(\boldsymbol{x})$ is also SOS for $t > t^*$.

The dual of a robust SOS-convex polynomial optimization problem previously appeared in Jeyakumar et al. (2015), where tractable problems were obtained for polyhedral and ellipsoidal uncertainty sets. It was shown that strong duality holds, but it remained unclear how to recover a primal optimal solution $\boldsymbol{x}$. With our results, any convex uncertainty set can be used and a primal optimal solution can be obtained.

4.4 Nonlinear in the optimization variables and nonlinear in the uncertain parameters

Separable. Let $\boldsymbol{h^a} : \mathbb{R}^L \to \mathbb{R}^k_+$ be a concave function, let $\boldsymbol{h^x} : \mathbb{R}^n \to \mathbb{R}^k_+$ be a convex function, and let $f(\boldsymbol{a},\boldsymbol{x}) = \boldsymbol{h^a}(\boldsymbol{a})^\top \boldsymbol{h^x}(\boldsymbol{x}) = \sum_{j=1}^k h^a_j(\boldsymbol{a}) h^x_j(\boldsymbol{x})$. The conjugate of $f(\boldsymbol{a},\boldsymbol{x})$ is:

$$f^*(\boldsymbol{a},\boldsymbol{u}) = \inf_{\boldsymbol{u^j}\in\mathbb{R}^n} \left\{ \sum_{j=1}^k h^a_j(\boldsymbol{a})(h^x_j)^* \left(\frac{\boldsymbol{u^j}}{h^a_j(\boldsymbol{a})} \right) : \sum_{j=1}^k \boldsymbol{u^j} = \boldsymbol{u} \right\}.$$

Second order cone. Let $\mathcal{U} \subset \mathbb{S}^n_{++} \times \mathbb{R}^n \times \mathbb{R}$, and let $f(\boldsymbol{A},\boldsymbol{c},d,\boldsymbol{x}) = \sqrt{(\boldsymbol{x}-\boldsymbol{b})^\top \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})} - \boldsymbol{c}^\top\boldsymbol{x} - d$. Indeed f models the conic quadratic constraint $\left|\left|\boldsymbol{A}^{\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{b})\right|\right|_2 \le \boldsymbol{c}^\top\boldsymbol{x} + d$, which is concave in the uncertain parameters and convex in $\boldsymbol{x}$. The convex conjugate of $f(\boldsymbol{A},\boldsymbol{c},d,\boldsymbol{x})$ is:

$$f^*(\boldsymbol{A},\boldsymbol{c},d,\boldsymbol{u}) = \begin{cases} \boldsymbol{b}^\top(\boldsymbol{u}+\boldsymbol{c}) + d & \text{if } \left|\left|\boldsymbol{A}^{-\frac{1}{2}}(\boldsymbol{u}+\boldsymbol{c})\right|\right|_2 \le 1 \\ \infty & \text{otherwise.} \end{cases}$$

The constraint $\left|\left|\boldsymbol{A}^{-\frac{1}{2}}(\boldsymbol{u}+\boldsymbol{c})\right|\right|_2 \le 1$ is not CQr, since $\boldsymbol{A}$, $\boldsymbol{u}$ and $\boldsymbol{c}$ are all variables. However, it is SDr via the Schur complement:

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{u}+\boldsymbol{c} \\ (\boldsymbol{u}+\boldsymbol{c})^\top & 1 \end{pmatrix} \in \mathbb{S}^{n+1}_+.$$

We have not found this conjugate in the literature. It can be verified that this formulation is also valid when $\boldsymbol{A}$ is positive *semi*definite.

Current known results for uncertain second order cone constraints are limited to $\boldsymbol{A} = \boldsymbol{B}^\top\boldsymbol{B}$ where $\boldsymbol{B}$ has interval uncertainty, norm-bounded uncertainty or ellipsoidal uncertainty (Ben-Tal et al. 2009, Ch. 6). We are the first to model uncertainty directly on $\boldsymbol{A}$, and the first to obtain results for any convex uncertainty set.

Negative square root. Let $\mathcal{U} \subset \mathbb{R}^n \times \mathbb{R}$, and let $f(\boldsymbol{a},c,\boldsymbol{X}) = -\sqrt{\boldsymbol{a}^\top\boldsymbol{X}\boldsymbol{a}} + c$ when $\boldsymbol{X} \in \mathbb{S}^n_+$, ∞ otherwise. This constraint is concave in the uncertain parameters and convex in $\boldsymbol{X}$. Let

$\delta(\boldsymbol{X}|\mathbb{S}^n_+)$ denote the indicator function on $\mathbb{S}^n_+$, taking the value 0 if $\boldsymbol{X} \in \mathbb{S}^n_+$ and ∞ otherwise. By definition, the convex conjugate of $f(\boldsymbol{a}, c, \boldsymbol{X})$ is:

$$\begin{aligned} f^*(\boldsymbol{a}, c, \boldsymbol{U}) &= \sup_{\boldsymbol{X} \in \mathbb{R}^{n \times n}} \{\mathrm{tr}(\boldsymbol{U}\boldsymbol{X}) + \sqrt{\boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{a}} - c - \delta(\boldsymbol{X}|\mathbb{S}^n_+)\} \\ &= \sup_{\boldsymbol{X} \in \mathbb{R}^{n \times n}} \{g(\boldsymbol{a}, c, \boldsymbol{U}, \boldsymbol{X}) - \delta(\boldsymbol{X}|\mathbb{S}^n_+)\}, \end{aligned}$$

with $g(\boldsymbol{a}, c, \boldsymbol{U}, \boldsymbol{X}) = \mathrm{tr}(\boldsymbol{U}\boldsymbol{X}) + \sqrt{\boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{a}} - c$, and where $\delta(\boldsymbol{X}|\mathbb{S}^n_+)$ takes the value 0 if $\boldsymbol{X}$ is positive semidefinite, ∞ otherwise. By Fenchel's duality theorem:

$$f^*(\boldsymbol{a}, c, \boldsymbol{U}) = \inf_{\boldsymbol{Z} \in \mathbb{R}^{n \times n}} \{\delta^*(\boldsymbol{Z}|\mathbb{S}^n_+) - g_*(\boldsymbol{a}, c, \boldsymbol{U}, \boldsymbol{Z})\}. \tag{16}$$

The conjugates can now easily be derived: $\delta^*(\boldsymbol{Z}|\mathbb{S}^n_+) = \delta(-\boldsymbol{Z}|\mathbb{S}^n_+)$, and $g_*(\boldsymbol{a}, c, \boldsymbol{U}, \boldsymbol{Z}) = c - \lambda/4$ when $\lambda \boldsymbol{Z} = \lambda \boldsymbol{U} + \boldsymbol{a}\boldsymbol{a}^\top$, ∞ otherwise. Plugging these into (16) and taking the Schur complement, we get:

$$f^*(\boldsymbol{a}, c, \boldsymbol{U}) = \inf_{\lambda} \left\{ \frac{1}{4}\lambda - c : \begin{pmatrix} -\boldsymbol{U} & \boldsymbol{a} \\ \boldsymbol{a}^\top & \lambda \end{pmatrix} \in \mathbb{S}^{n+1}_+ \right\}.$$

Exponential function. Let $\mathcal{U} \subset \mathbb{R}^n_{++}$ and let $f(\boldsymbol{a}, \boldsymbol{x}) = \sum_{j=1}^n (a_j)^{x_j}$ if $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{1}$, $f(\boldsymbol{a}, \boldsymbol{x}) = \infty$ otherwise. Since $f(\boldsymbol{a}, \boldsymbol{x})$ is separable into functions $f_j(a_j, x_j) = (a_j)^{x_j}$ if $0 \leq x_j \leq 1$, ∞ otherwise, the conjugate of $f^*(\boldsymbol{a}, \boldsymbol{u})$ is the sum of the conjugates: $\sum_{j=1}^n f_j^*(a_j, u_j)$, each given by:

$$f_j^*(a_j, u_j) = \begin{cases} -1 & \text{if } u_j \leq \log(a_j) \\ \frac{u_j}{\log a_j} \log\left(\frac{u_j}{\log a_j}\right) - \frac{u_j}{\log a_j} & \text{if } \log(a_j) < u_j < a_j \log(a_j) \\ u_j - a_j & \text{if } a_j \log(a_j) \leq u_j. \end{cases}$$

This function is convex and continuous, but not differentiable. This may lead to numerical issues, depending on the solver.

Power function. Let $\mathcal{U} \subset [0,1]^n$ and let $f(\boldsymbol{a}, \boldsymbol{x}) = -\sum_{j=1}^n x_j^{a_j}$ if $\boldsymbol{x} > \boldsymbol{0}$, $f(\boldsymbol{a}, \boldsymbol{x}) = \infty$ otherwise. Since $f(\boldsymbol{a}, \boldsymbol{x})$ is separable into functions $f_j(a_j, x_j) = -x_j^{a_j}$ if $x_j > 0$, ∞ otherwise, the conjugate $f^*(\boldsymbol{a}, \boldsymbol{u})$ is the sum of the conjugates: $\sum_{j=1}^n f_j^*(a_j, u_j)$, each given by:

$$f_j^*(a_j, u_j) = \begin{cases} 1 & \text{if } a_j = 0 \text{ and } u_j \leq 0 \\ 0 & \text{if } a_j = 1 \text{ and } u_j = -1 \\ \left(-\frac{u_j}{a_j}\right)^{\frac{a_j}{a_j - 1}} (1 - a_j) & \text{if } 0 < a_j < 1 \text{ and } u_j \leq 0 \\ \infty & \text{else.} \end{cases}$$

This function is convex and continuous, but not differentiable. However, an interior point method can use the third formulation on the interior of the feasible region, and converges to the global optimum.

4.5 Globalized Robust Counterpart

In order to reduce the conservatism of the RC, Ben-Tal et al. (2006) propose to use a small uncertainty set of parameters for which a constraint has to hold, and a second, larger uncertainty set for which the constraint may be violated to some degree. The allowable violation depends on the distance between the realized parameter and the smaller set. Let $\mathcal{U}'$ denote the smaller set, and let $\mathcal{U} \supset \mathcal{U}'$ denote the larger set, then the *Globalized Robust Counterpart* (GRC) is given by:

$$g(\boldsymbol{a}, \boldsymbol{x}) \leq \inf_{\boldsymbol{a}' \in \mathcal{U}'} \{h(\boldsymbol{a}, \boldsymbol{a}')\} \qquad \forall \boldsymbol{a} \in \mathcal{U}, \tag{17}$$

where $h(\boldsymbol{a}, \boldsymbol{a}')$ is a nonnegative jointly convex distance-like function for which $h(\boldsymbol{a}', \boldsymbol{a}') = 0$ for all $\boldsymbol{a}'$ in $\mathcal{U}'$. To put this constraint in the general framework, define:

$$f(\boldsymbol{a}, \boldsymbol{x}) = g(\boldsymbol{a}, \boldsymbol{x}) - \inf_{\boldsymbol{a}' \in \mathcal{U}'} \{h(\boldsymbol{a}, \boldsymbol{a}')\}.$$

Note that indeed $f(\boldsymbol{a}, \boldsymbol{x})$ is concave in $\boldsymbol{a}$ (Boyd and Vandenberghe 2004, §3.2.5). The GRC (17) was introduced by Gorissen et al. (2014), who rewrite it to an equivalent constraint with finitely many variables and constraints using the concave conjugate of $f(\boldsymbol{a}, \boldsymbol{x})$ and the convex conjugate of $h(\boldsymbol{a}, \boldsymbol{a}')$ (both w.r.t. $\boldsymbol{a}$) and the support function of $\mathcal{U}$. Gorissen et al. (2014) show how to solve the GRC via the dual if $g(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x}$. We generalize the result by Gorissen et al. (2014) to any nonlinear concave-convex function $g(\boldsymbol{a}, \boldsymbol{x})$ for which the convex conjugate exists. Since the right hand side of (17) does not depend on $\boldsymbol{x}$, it does not complicate the derivation of the conjugate:

$$f^*(\boldsymbol{a},\boldsymbol{u}) = g^*(\boldsymbol{a},\boldsymbol{u}) + \inf_{\boldsymbol{a'}\in\mathcal{U}'}\{h(\boldsymbol{a},\boldsymbol{a'})\}.$$

This formula can directly be used for (COD) since the min operator may be omitted, so that $\boldsymbol{a'}$ becomes an optimization variable of (COD). Therefore, if the RC $g(\boldsymbol{a},\boldsymbol{x}) \leq 0$ for all $\boldsymbol{a}$ in $\mathcal{U}$ is tractable with our method, then so is the GRC.

4.6 Illustrative example to formulate the dual problem

In the previous section, we derived many conjugate functions. We now demonstrate how these can be used to formulate the dual problem (COD). As an illustrative example, we consider the following problem:

$$\begin{aligned}
\text{(Example)} \quad \sup_{\boldsymbol{x}\in\mathbb{R}^n_+} \quad & \boldsymbol{c}^\top\boldsymbol{x} \\
\text{s.t.} \quad & \sqrt{(\boldsymbol{x}-\boldsymbol{b})^\top\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})} \leq \boldsymbol{c}^\top\boldsymbol{x} + d \quad \forall \boldsymbol{A} : \sum_{i=1}^n\sum_{j=1}^n(\boldsymbol{A}_{ij} - \bar{\boldsymbol{A}}_{ij})^2 \leq \rho \\
& \boldsymbol{x}^\top\boldsymbol{B}\boldsymbol{x} + \boldsymbol{b}^\top\boldsymbol{x} + c \leq 0 \quad \forall \boldsymbol{B}\in\mathbb{S}^n_+ : \boldsymbol{I}-\boldsymbol{B}\in\mathbb{S}^n_+ \\
& 1 - \sum_{i=1}^n |1 - a_i^{\frac{1}{3}}|^3 x_i \leq 0 \quad \forall \boldsymbol{a}\in[0,1]^n : \left\|\boldsymbol{a} - \frac{1}{2}\boldsymbol{e}\right\|_2 \leq \rho.
\end{aligned}$$

Let $f_1(\boldsymbol{A},\boldsymbol{x})$, $f_2(\boldsymbol{B},\boldsymbol{x})$ and $f_3(\boldsymbol{a},\boldsymbol{x})$ denote the constraint functions. The conjugates for these constraints are given in Sections 4.4, 4.2, and 4.1, respectively. After fixing the parameters that are not uncertain, we obtain:

$$y_1 f_1^*\left(\frac{\boldsymbol{V^1}}{y_1}, \frac{\boldsymbol{w^1}}{y_1}\right) = \begin{cases} \boldsymbol{b}^\top(\boldsymbol{w^1} + y_1\boldsymbol{c}) + y_1 d & \text{if } \begin{pmatrix} \boldsymbol{V^1} & \boldsymbol{w^1} + y_1\boldsymbol{c} \\ (\boldsymbol{w^1} + y_1\boldsymbol{c})^\top & y_1 \end{pmatrix} \in \mathbb{S}^{n+1}_+ \\ \infty & \text{otherwise,} \end{cases}$$

while $y_2 f_2^*\left(\frac{\boldsymbol{V^2}}{y_2}, \frac{\boldsymbol{w^2}}{y_2}\right) \leq t$ is equivalent to:

$$\begin{pmatrix} 4\boldsymbol{V^2} & \boldsymbol{w^2} - y_2\boldsymbol{b} \\ (\boldsymbol{w^2} - y_2\boldsymbol{b})^\top & c y_2 + t \end{pmatrix} \in \mathbb{S}^{n+1}_+,$$

and

$$
y_3 f_3^* \left(\frac{\boldsymbol{v^3}}{y_3}, \frac{\boldsymbol{w^3}}{y_3} \right) = \begin{cases} -y_3 & \text{if } \boldsymbol{w_i^3} \leq -|y_3^{\frac{1}{3}} - (\boldsymbol{v_i^3})^{\frac{1}{3}}|^3 \\ \infty & \text{otherwise.} \end{cases}
$$

We explicitly show how to formulate the perspective of the constraint that defines the first uncertainty set. The first uncertainty set is defined by the function $g_{11} : \mathbb{R}^{n \times n} \to \mathbb{R}$, given by:

$$
g_{11}(\boldsymbol{A}) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\boldsymbol{A}_{ij} - \bar{\boldsymbol{A}}_{ij})^2 - \rho,
$$

so, constraint (8), $y_1 g_{11}(\boldsymbol{V^1}/y_1) \leq 0$, is equivalent to:

$$
y_1 \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\boldsymbol{V_{ij}^1}}{y_1} - \bar{\boldsymbol{A}}_{ij} \right)^2 \leq \rho y_1.
$$

Multiplying both sides with y_1 yields:

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\boldsymbol{V_{ij}^1} - \bar{\boldsymbol{A}}_{ij} y_1 \right)^2 \leq \rho y_1^2,
$$

which is CQr. Combining all results into (7), we get:

$$
\begin{aligned}
\text{(COD-Example)} \quad \inf \quad & \boldsymbol{b}^\top(\boldsymbol{w^1} + y_1\boldsymbol{c}) + y_1 d + t - y_3 \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{V^1} & \boldsymbol{w^1} + y_1\boldsymbol{c} \\ (\boldsymbol{w^1} + y_1\boldsymbol{c})^\top & y_1 \end{pmatrix} \in \mathbb{S}_+^{n+1} \\
& \begin{pmatrix} 4\boldsymbol{V^2} & \boldsymbol{w^2} - y_2\boldsymbol{b} \\ (\boldsymbol{w^2} - y_2\boldsymbol{b})^\top & cy_2 + t \end{pmatrix} \in \mathbb{S}_+^{n+1} \\
& \boldsymbol{w}_i^{\boldsymbol{3}} \leq -|y_3^{\frac{1}{3}} - (\boldsymbol{v}_i^{\boldsymbol{3}})^{\frac{1}{3}}|^3 \quad \forall i = 1, \dots, n \\
& \sum_{i=1}^n \boldsymbol{w^i} = \boldsymbol{c} \qquad (18) \\
& \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{V}_{ij}^{\boldsymbol{1}} - \bar{\boldsymbol{A}}_{ij} y_1)^2 \leq \rho y_1^2 \\
& y_2 \boldsymbol{I} - \boldsymbol{V^2} \in \mathbb{S}_+^n \\
& \boldsymbol{v}_i^{\boldsymbol{3}} \leq y_3 \quad \forall i = 1, \dots, n \\
& \left\| \boldsymbol{v^3} - \frac{1}{2} y_3 \boldsymbol{e} \right\|_2 \leq \rho y_3 \\
& \boldsymbol{y} \in \mathbb{R}_+^3,\ \boldsymbol{V^1} \in \mathbb{R}^{n \times n},\ \boldsymbol{V^2} \in \mathbb{S}_+^n,\ \boldsymbol{v^3} \in \mathbb{R}_+^n,\ \boldsymbol{w^i} \in \mathbb{R}^n.
\end{aligned}
$$

An optimal solution $\boldsymbol{x}$ to (Example) is now given by the dual variable with respect to constraint (18) in an optimal solution of (COD-Example).

5 Concluding remark

Our method can systematically solve nonlinear RO problems, and complements the primal method by Ben-Tal et al. (2015). Currently, software for automatic robust optimization is focused on the primal problem (Dunning 2016, Goh and Sim 2011, Löfberg 2012, Roelofs and Bisschop 2015), and the user is therefore limited to a few choices for the uncertainty set. Our method can extend these automated methods to nonlinear optimization and to a large class of uncertainty sets.

Acknowledgment

The authors thank F. J. C. T. de Ruiter (Tilburg University) for his comments on a previous version of this paper.

References

A. Ahmadi and P. Parrilo. A complete characterization of the gap between convexity and SOS-convexity. *SIAM Journal on Optimization*, 23(2):811–833, 2013.

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

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

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

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

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

D. Bertsimas, I. Dunning, and M. Lubin. Reformulations versus cutting planes for robust optimization: A computational and machine learning perspective. *Optimization Online*, November, 2014.

S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

I. Dunning. JuMPeR – robust optimization with JuMP, 2016. http://iainnz.github.io/JuMPeR.jl/, retrieved February 2016.

J. Goh and M. Sim. Robust optimization made easy with ROME. *Operations Research*, 59(4):973–985, 2011.

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

B. L. Gorissen, A. Ben-Tal, J. P. C. Blanc, and D. den Hertog. Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Operations Research*, 62(3):672–679, 2014.

J. W. Helton and J. Nie. Semidefinite representation of convex sets. *Mathematical Programming*, 122 (1):21–64, 2010.

V. Jeyakumar, G. Li, and J. Vicente-Pérez. Robust SOS-convex polynomial optimization problems: exact SDP relaxations. *Optimization Letters*, 9(1):1–18, 2015.

M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In M. Putinar and S. Sullivant, editors, *Emerging Applications of Algebraic Geometry*, volume 149 of *The IMA Volumes in Mathematics and its Applications*. Springer New York, 2009.

J. Löfberg. Automatic robust convex programming. *Optimization Methods and Software*, 27(1):115–129, 2012.

S. Mehrotra and D. Papp. A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. *SIAM Journal on Optimization*, 24(4):1670–1697, 2014.

Y. Nesterov and A. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming.* SIAM, Princeton, NJ, 1994.

K. Postek, D. den Hertog, and B. Melenberg. Tractable counterparts of distributionally robust constraints on risk measures. *CentER Discussion Paper*, 2014-031, 2014. To appear in Siam Review.

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

M. Roelofs and J. Bisschop. AIMMS 4.3: The Language Reference, Paragon Decision Technology BV, 2015.

A. Y. D. Siem, E. de Klerk, and D. den Hertog. Discrete least-norm approximation by nonnegative (trigonometric) polynomials and rational functions. *Structural and Multidisciplinary Optimization*, 35(4):327–339, 2008.