Hidden convexity in a class of optimization problems with bilinear terms

Bram L. Gorissen[1], Dick den Hertog[2], Meike Reusken[3]

Abstract

In this paper we identify a new class of nonconvex optimization problems that can be equivalently reformulated to convex ones. These nonconvex problems can be characterized by convex functions with bilinear arguments. We describe several examples of important applications that have this structure. A reformulation technique is presented which converts the problems in this class into convex and tractable problems. We show that this reformulation technique can be used to develop the dual of robust nonlinear optimization problems that are convex in the optimization variables and concave in the uncertain parameters. To find the dual of such problems we employ the 'primal worst equals dual best' technique, where the uncertain parameters become variables in the dual. We show that the 'dual best' formulation has the hidden convexity structure studied in this paper, and therefore can be reformulated into a tractable convex optimization problem. Additionally, we show that inverse optimization problems for general linear and nonlinear optimization problems also have the hidden convexity structure, and hence can also be reformulated as convex problems. The value of the reformulation is illustrated by several numerical experiments for the nutritious food supply chain model for the World Food Programme.

Keywords: nonconvex optimization; hidden convexity; robust optimization; inverse optimization

MSC classification: 90C30

1 Introduction

Convexity is a crucial property in optimization, since for convex problems it holds that every local optimum is also a global optimum. Some nonconvex problems admit an equivalent convex

[1]Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School. E-mail: bgorissen@broadinstitute.org

[2]Amsterdam Business School, University of Amsterdam.

[3]Department of Econometrics and Operations Research, Tilburg University. E-mail: m.c.d.reusken_1@tilburguniversity.edu

[4]This paper includes our previous working paper "Robust nonlinear optimization via the dual" that has not been and will not be published elsewhere.

reformulation, and such problems are called hidden convex problems. In the literature several classes of problems with hidden convexity have been detected (Ben-Tal and Teboulle 1996, Wu et al. 2007, Ben-Tal et al. 2011, Rockafellar 2022). In this paper we identify a new class of nonconvex optimization problems with hidden convexity. These nonconvex problems contain convex objective and constraint functions with bilinear arguments. We describe three classes of important applications that fit into this structure:

1. **Problems with variable coefficients.** The first class consists of problems with variable columns of coefficients. We will give two toy examples to illustrate the structure of such problems. We let $\mathbf{1}$ denote the vector of all 1s, with dimension implied by the context. The first example

$$\text{(T1)} \quad \min_{\boldsymbol{x}\in\mathbb{R}^2_+, \boldsymbol{z}\in\mathbb{R}^4} \quad -3x_1 - 4x_2$$

$$\text{s.t.} \quad \begin{pmatrix} 1 + z_1 + 2z_2 \\ z_1 \\ 2 + z_1 + z_2 \end{pmatrix} x_1 + \begin{pmatrix} 1 + 2z_3 \\ z_3 + 3z_4 \\ 2 \end{pmatrix} x_2 \leq \mathbf{1}$$

$$z_1^2 + z_2^2 \leq 1 \tag{1}$$

$$z_3^2 + z_4^2 \leq 1 \tag{2}$$

$$\boldsymbol{x} \leq \mathbf{1},$$

has bilinear terms and the variables that appear in the coefficients for x_1 (z_1 and z_2) do not appear in the coefficients for x_2, and the coefficients for x_2 (z_3 and z_4) do not appear in the coefficients for x_1. Moreover, the constraints on z are 'columnwise', i.e., the constraints (1) and (2) are separate in (z_1, z_2) and (z_3, z_4). Similar classes of problems were considered by Dantzig (1963). He called such problems Generalized Linear Programming (GLP) problems. He showed 'hidden convexity' for linear optimization problems with variable coefficients, and where the constraints on the variables in a column are linear. We extend these results for GLP in two ways. First of all, our reformulation allows the constraints on z to be nonlinear. Moreover, we show hidden convexity for *nonlinear* problems with variable coefficients. The structure of such problems is illustrated in the next toy example:

$$
\begin{aligned}
\text{(T2)} \quad \min_{\boldsymbol{x}\in\mathbb{R}^2_+,\boldsymbol{z}\in\mathbb{R}^4} \quad & 3x_1 + 4x_2 \\
\text{s.t.} \quad & e^{(1+z_1+2z_2)x_1} + e^{(1+2z_3)x_2} \leq 2 \\
& e^{z_1^2 x_1} + e^{(z_3+e^{3z_4})x_2} \leq 2 \\
& e^{(2+e^{-z_1}+z_2)x_1} + e^{2x_2} \leq 2 \\
& z_1^2 + z_2^2 \leq 2 \\
& z_3^2 + z_4^2 \leq 2 \\
& \boldsymbol{x} \leq \mathbf{1}.
\end{aligned}
$$

The bilinear structure in this problem is maybe not obvious. However, we could replace the nonlinear coefficient z_1^2 by y_1, and add the constraint $y_1 \geq z_1^2$, do the same with the two other nonlinear coefficients, and then the bilinear structure appears. Moreover, again the variables in the coefficients are columnwise, i.e., the variables z_1 and z_2 only occur in the coefficients for x_1. The constraints on $\boldsymbol{z}$ are separate for (z_1, z_2) and (z_3, z_4).

In this paper we show how we can reformulate nonconvex problems, such as (T1) and (T2), into convex ones. We also apply these techniques to two such optimization problems for the food supply chain of the UN World Food Programme.

2. **The dual of a robust nonlinear optimization problem.** 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, Bertsimas and den Hertog 2022). In basic versions of RO, the constraints have to hold for all parameter realizations in a prespecified (infinite) uncertainty set. 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. The most generic results for this dualization approach are by Ben-Tal et al. (2015).

   In Beck and Ben-Tal (2009) the dual of the robust nonlinear problem is studied. That paper shows that in the dual the $\forall$ quantifier for the uncertain parameter becomes a $\exists$

quantifier, i.e., in the dual the uncertain parameters become variables. This is summarized as 'primal worst equals dual best'. This result of the paper is used in more than 180 papers, but in all these papers the 'dual best' problem remains nonconvex, and is therefore not useful for computational purposes. In this paper we show that the 'dual best' problem has the hidden convexity structure studied in this paper, and therefore can be reformulated into a convex optimization problem. It extends previous results on robust *linear* optimization (Gorissen et al. 2014) to the nonlinear case. We derive 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. This new approach of solving the dual offers several advantages. It may be easier to solve or formulate the dual, and the dual can be formulated for any convex uncertainty set.

3. **Inverse optimization.** Inverse optimization is about recovering optimization parameters from an optimal solution. Approaches to find the objective coefficients of a linear optimization problem were first studied by Ahuja and Orlin (2001). Using the complementary slackness optimality condition, they describe a linear inverse optimization model for this setting. Subsequent research remained focused on estimating the linear objective, but investigated different optimality conditions (Shahmoradi and Lee 2021, Chan et al. 2018). Inverse optimization approaches for problems with conic constraints and certain types of quadratic objectives are proposed in Iyengar and Kang (2005). A different approach for problems with quadratic objectives and linear constraints is proposed in Zhang and Zhang (2010). Zhang and Xu (2010) propose an inverse optimization method for optimization problems with a separable objective function and up to one separable constraint. Chan and Kaw (2020) add to this literature by imputing left-hand-side constraint coefficients, in addition to objective coefficients. Using the hidden convexity reformulation technique introduced in this paper, we extend the state-of-the-art results in inverse optimization in two ways. Firstly, in case the inverse optimization problem is to recover coefficients in both the linear objective and the constraints, our approach admits a set of admissible objective coefficients, contrary to Chan and Kaw (2020). Secondly, our approach can also be used for a broad class of nonlinear optimization problems, contrary to the existing literature. In this paper we also show how our approach can be used for an inverse optimization problem for the World Food Programme.

We summarize the contributions of this paper:

1. We prove hidden convexity for a new rich class of nonconvex optimization problems. We show that these problems can be equivalently reformulated as convex ones.

2. We show that the dual of a robust nonlinear optimization problem falls into this class of problems with hidden convexity. Hence, as we will show, we are able to reformulate the dual problem to a convex one. As a side result, we derive computationally tractable robust counterparts for several new robust nonlinear optimization problems.

3. We show that inverse optimization problems for general linear and nonlinear optimization problems also have the hidden convexity structure, and hence can also be reformulated as convex problems. This significantly extends the existing approaches in the literature on inverse optimization to a broad class of nonlinear problems.

4. Three applications for the food supply chain of the UN World Food Programme show the value of our reformulation.

Notations.

We use non-boldface characters (x) to denote scalars, lowercase boldface characters ($\boldsymbol{x}$) to denote vectors, uppercase boldface characters ($\boldsymbol{X}$) to denote matrices.

We use $[n]$ to denote the finite index set $[n] = \{1, \ldots, n\}$ with cardinality $|[n]| = n$.

The function $\delta^*(\boldsymbol{x}|\boldsymbol{S})$ denotes the support function of the set $\boldsymbol{S}$ evaluated at $\boldsymbol{x}$, i.e., $\delta^*(\boldsymbol{x}|\boldsymbol{S}) = \sup_{\boldsymbol{y}\in\boldsymbol{S}} \boldsymbol{y}^T\boldsymbol{x}$.

A convex function $f : \mathbb{R}^{n_y} \to [-\infty, +\infty]$ is said to be *proper* if there exists at least one vector $\boldsymbol{y}$ such that $f(\boldsymbol{y}) < +\infty$ and for all $\boldsymbol{y} \in \mathbb{R}^{n_y}$, $f(\boldsymbol{y}) > -\infty$. For a proper convex function f, we define its *convex conjugate* as $f^*(\boldsymbol{w}) = \sup_{\boldsymbol{y}\in\operatorname{dom} f} \left\{\boldsymbol{w}^T\boldsymbol{y} - f(\boldsymbol{y})\right\}$, where $\operatorname{dom} f$ is the effective domain: $\{\boldsymbol{y} : f(\boldsymbol{y}) < +\infty\}$. When f is closed and convex, $f^{**} = f$. The *concave conjugate* of a function g is $g_*(\boldsymbol{w}) = \inf_{\boldsymbol{y}\in\operatorname{dom} g} \left\{\boldsymbol{w}^T\boldsymbol{y} - g(\boldsymbol{y})\right\}$. 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 $g : \mathbb{R}^{n_x} \to [-\infty, +\infty]$ is defined by $h : \mathbb{R}^{n_x} \times \mathbb{R}_+$, $h(\boldsymbol{x}, y) = yg(\boldsymbol{x}/y)$ for $y > 0$ and $h(\boldsymbol{x}, 0) = \lim_{y\downarrow 0} yg(\boldsymbol{x}/y)$. For brevity and clarity, we will not introduce specific functions for each perspective. Instead, we denote the perspective of g as $yg(\boldsymbol{x}/y)$, with the convention that for $y = 0$, the function value is defined as a limit. The perspective of a convex function is convex (Rockafellar 1970, p. 35).

2 General Framework

We will show hidden convexity for the following class of problems:

$$\text{(HC)} \quad \min_{\boldsymbol{x}\in\mathbb{R}^n_+,\ \boldsymbol{z^k}\in\mathbb{R}^{r_k}} \quad f_0\left(\sum_{k\in[n]} \boldsymbol{a^{0k}}\left(\boldsymbol{z^k}\right)x_k + \boldsymbol{b^0}\left(\boldsymbol{z^{n+1}}\right)\right) \tag{3}$$

$$\text{s.t.} \quad f_i\left(\sum_{k\in[n]} \boldsymbol{a^{ik}}\left(\boldsymbol{z^k}\right)x_k + \boldsymbol{b^i}\left(\boldsymbol{z^{n+1}}\right)\right) \leq 0 \qquad \forall i \in [I] \tag{4}$$

$$g_{kl}(\boldsymbol{z^k}) \leq 0 \qquad \forall k \in [n+1] \quad \forall l \in [L_k] \tag{5}$$

$$h_j(\boldsymbol{x}) \leq 0 \qquad \forall j \in [J],$$

where $\boldsymbol{a^{ik}} : \mathbb{R}^{r_k} \to \mathbb{R}^m$, $\boldsymbol{b^i} : \mathbb{R}^{r_{n+1}} \to \mathbb{R}^m$, $g_{kl} : \mathbb{R}^{r_k} \to \mathbb{R}$, $h_j : \mathbb{R}^n \to \mathbb{R}$, and $f_i : \mathbb{R}^m \to \mathbb{R}$ are convex. Moreover, for each i in $\{0\} \cup [I]$, either (i) $\boldsymbol{a^{ik}}$ and $\boldsymbol{b^i}$ are affine, or (ii) $f_i(\boldsymbol{y})$ is nondecreasing in y_j, $\forall j \in [m]$. Hence, this class of problems has convex problem functions, but their arguments are bilinear in the optimization variables $\boldsymbol{x}$ and $\boldsymbol{z}$ in case $\boldsymbol{a^{ik}}$ is affine, or bilinear in $\boldsymbol{x}$ and $\boldsymbol{a^{ik}}(\boldsymbol{z^k})$ otherwise. Therefore, the problem is nonconvex in the optimization variables. The problem also has a set of convex constraints on $\boldsymbol{x}$, and a set of convex constraints on $\boldsymbol{z^k}$ only. Hence, this problem has a 'columnwise structure', i.e., the column $\boldsymbol{a^{0k}}, \cdots, \boldsymbol{a^{Ik}}$ only depends on $\boldsymbol{z^k}$, and the constraints (5) are separable in $\boldsymbol{z^k}$. To illustrate the structure of problem (HC) we write the functions $\boldsymbol{a^{i1}}, \cdots, \boldsymbol{a^{in}}$ as a matrix valued function $\boldsymbol{A^i}$, and then write the objective function (3) and constraint functions in (4) as $f_i(\boldsymbol{A^i}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{b^i}(\boldsymbol{z}))$.

To illustrate the richness of this class of problems, we describe two examples of classes of optimization problems that fit into this class. In the next two sections we will show that the dual of robust nonlinear problems and inverse nonlinear optimization also fit in this structure.

Example 1 *The problem class (HC) contains Generalized Linear Programming (GLP) problems studied by Dantzig (1963), in which the functions f_i, $\boldsymbol{a^{ik}}$, $\boldsymbol{b^i}$, g_{kl}, and h_j, $i \in \{0\} \cup [I]$, $j \in [J]$, $k \in [n+1]$, $l \in [L_k]$, are linear.*

Example 2 *The problem class (HC) contains quadratic optimization with variable coefficients:*

$$
\begin{aligned}
\text{(GQP)} \quad \min_{\boldsymbol{x}\in\mathbb{R}^n_+, \boldsymbol{z^k}\in\mathbb{R}^m} \quad & \boldsymbol{x}^T\boldsymbol{A^0}(\boldsymbol{z})^T\boldsymbol{A^0}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{b^0}(\boldsymbol{z})^T\boldsymbol{x} + c^0 \\
\text{s.t.} \quad & \boldsymbol{x}^T\boldsymbol{A^i}(\boldsymbol{z})^T\boldsymbol{A^i}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{b^i}(\boldsymbol{z})^T\boldsymbol{x} + c^i \le 0 \qquad \forall i \in [m] \\
& g_{kl}(\boldsymbol{z^k}) \le 0 \qquad \forall k \in [n+1] \quad \forall l \in [L_k] \\
& h_j(\boldsymbol{x}) \le 0 \qquad \forall j \in [J],
\end{aligned}
$$

where $\boldsymbol{A^i(z)}$ has a column-wise structure, i.e., its k-th column $\boldsymbol{a^{ik}(z)}$ depends only on $\boldsymbol{z^k}$, and $b^{ik}(\boldsymbol{z})$ depends only on $\boldsymbol{z^k}$. Moreover, in this case $\boldsymbol{a^{ik}(z^k)}$ and $\boldsymbol{b^{ik}(z^k)}$ need to be affine in $\boldsymbol{z^k}$, since the quadratic function is not nondecreasing.

Examples of functions $f_i(\cdot)$ that are convex and nondecreasing are the sum of maxima of linear functions and the log-sum-exp function.

We now show that problem (HC) can be equivalently reformulated into a convex one.

Theorem 1 *Assume that for each $k \in [n]$, either $x_k > 0$ (i.e., $\nexists\, \boldsymbol{x} \in \mathbb{R}^n_+ \ : \ x_k = 0 \ \wedge \ h_j(\boldsymbol{x}) \le 0 \ \forall j \in [J]$) or the set $S = \{\boldsymbol{z^k} : g_{kl}(\boldsymbol{z^k}) \le 0 \quad \forall l \in [L_k]\}$ is bounded, and that for each i in $\{0\} \cup [I]$, either (i) $\boldsymbol{a^{ik}}$ and $\boldsymbol{b^i}$ are affine, or (ii) $f_i(\boldsymbol{y})$ is nondecreasing in y_j, $\forall j \in [m]$, then (HC) is equivalent to the following convex problem:*

$$
\begin{aligned}
(C) \quad \min_{\boldsymbol{x}\in\mathbb{R}^n_+, \boldsymbol{y^k}\in\mathbb{R}^{r_k}, \boldsymbol{z^{n+1}}\in\mathbb{R}^{r_{n+1}}} \quad & f_0\left(\sum_{k\in[n]} x_k\boldsymbol{a^{0k}}\left(\frac{\boldsymbol{y^k}}{x_k}\right) + \boldsymbol{b^0}\left(\boldsymbol{z^{n+1}}\right)\right) \\
s.t. \quad & f_i\left(\sum_{k\in[n]} x_k\boldsymbol{a^{ik}}\left(\frac{\boldsymbol{y^k}}{x_k}\right) + \boldsymbol{b^i}\left(\boldsymbol{z^{n+1}}\right)\right) \le 0 \qquad \forall i \in [I] \qquad (6) \\
& x_k g_{kl}\left(\frac{\boldsymbol{y^k}}{x_k}\right) \le 0 \qquad \forall k \in [n] \quad \forall l \in [L_k] \\
& g_{(n+1)l}(\boldsymbol{z^{n+1}}) \le 0 \quad \forall l \in [L_{n+1}] \\
& h_j(\boldsymbol{x}) \le 0 \qquad \forall j \in [J].
\end{aligned}
$$

Proof. To show that problem (C) is convex, we distinguish between two cases: (i) if $\boldsymbol{a^{ik}}$ and $\boldsymbol{b^i}$ are affine, then $x_k\boldsymbol{a^{ik}}\left(\boldsymbol{y^k}/x_k\right) + \boldsymbol{b^i}\left(\boldsymbol{z^{n+1}}\right)$ is affine in $(x_k, \boldsymbol{y^k}, \boldsymbol{z^{n+1}})$, and hence the objective and constraints (6) are convex, and (ii) if $f_i(\boldsymbol{y})$ is nondecreasing in y_j, $\forall j \in [m]$, then the objective and constraints (6) are also convex, since $x_k\boldsymbol{a^{ik}}\left(\boldsymbol{y^k}/x_k\right) + \boldsymbol{b^i}\left(\boldsymbol{z^{n+1}}\right)$ is convex in

$(x_k, \boldsymbol{y^k}, \boldsymbol{z^{n+1}})$.

To show that (HC) and (C) are equivalent, we claim that their feasible points are equivalent via the relation $\boldsymbol{y^k} = \boldsymbol{z^k} x_k$ for $k \in [n]$ and that corresponding points have the same objective value. If $x_k > 0$, the equivalence is trivial. If $x_k = 0$ and S is bounded, $\boldsymbol{y^k} = \boldsymbol{0}$ (Gorissen et al. 2014, Lemma 1), so the term $x_k \boldsymbol{a^{ik}} \left(\boldsymbol{y^k}/x_k\right)$ in the objective and constraint functions of (HC) equals $\lim_{x_k \downarrow 0} x_k \boldsymbol{a^{ik}}(\boldsymbol{0}) = \boldsymbol{0}$, which is indeed equal to the corresponding value of the term $\boldsymbol{a^{ik}} \left(\boldsymbol{z^k}\right) x_k$ in (C) irrespective of the value of $\boldsymbol{z^k}$. ■

3 The Dual of Robust Nonlinear Optimization Problems

In this section we show that the dual of a Robust Nonlinear Optimization problem also contains hidden convexity of the same type as described in the previous section. Hence, we are able to derive a convex reformulation of the dual problem, even for problems for which there is no computationally tractable formulation of the primal problem.

3.1 The primal robust counterpart

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

$$\begin{aligned} \text{(RC)} \quad & \max_{\boldsymbol{x} \in \mathbb{R}^n} \quad \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{s.t.} \quad f_i(\boldsymbol{a^i}, \boldsymbol{x}) \leq 0 \qquad \forall \boldsymbol{a^i} \in \mathcal{U}_i \quad \forall i \in [I]. \end{aligned} \tag{7}$$

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 (7) 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:

$$(\boldsymbol{a^i})^\top \boldsymbol{x} \leq b_i \;\forall \boldsymbol{a^i} : \boldsymbol{D}\boldsymbol{a^i} \leq \boldsymbol{d} \;\Leftrightarrow\; \sup_{\boldsymbol{a^i}:\boldsymbol{D}\boldsymbol{a^i}\leq \boldsymbol{d}} (\boldsymbol{a^i})^\top \boldsymbol{x} \leq b_i \;\Leftrightarrow\; \inf_{\boldsymbol{y^i}\geq \boldsymbol{0}:\boldsymbol{D}^\top \boldsymbol{y^i}=\boldsymbol{x}} \boldsymbol{d}^\top \boldsymbol{y^i} \leq b_i. \tag{8}$$

The inf operator can now be omitted, since if the inequality holds for one $\boldsymbol{y^i}$, it holds for the infimum. The most generic results for this dualization approach are by Ben-Tal et al. (2015).

Beck and Ben-Tal (2009) studied the dual of the robust nonlinear problem (7). They show that in the dual the $\forall$ quantifier for the uncertain parameter becomes a $\exists$ quantifier, i.e., in the dual the uncertain parameters become variables. They summarize this as 'primal worst equals dual best'. Their result is used in more than 180 papers, but for these papers the 'dual best' problem remains nonconvex, and is therefore not useful for computational purposes. In this paper we show that the 'dual best' problem also has the hidden convexity structure studied in this paper, and therefore can be reformulated into a convex optimization problem.

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:

$$\begin{aligned}
(f_i)_*(\boldsymbol{v^i}, \boldsymbol{x}) &= \inf_{\boldsymbol{a^i}\in\mathbb{R}^L} \{(\boldsymbol{v^i})^\top \boldsymbol{a^i} - f_i(\boldsymbol{a^i}, \boldsymbol{x})\},\\
f_i^*(\boldsymbol{a^i}, \boldsymbol{u^i}) &= \sup_{\boldsymbol{x}\in\mathbb{R}^n} \{(\boldsymbol{u^i})^\top \boldsymbol{x} - f_i(\boldsymbol{a^i}, \boldsymbol{x})\}, \text{ and}\\
g_{ik}^*(\boldsymbol{v^i}) &= \sup_{\boldsymbol{a^i}\in\mathbb{R}^L} \{(\boldsymbol{v^i})^\top \boldsymbol{a^i} - g_{ik}(\boldsymbol{a^i})\}.
\end{aligned}$$

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.

3.2 Hidden convexity in the dual

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 $\max_{\boldsymbol{x}\in\mathbb{R}^n}\{\boldsymbol{c}^\top \boldsymbol{x}$:

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

$$\inf_{\boldsymbol{y}\geq \boldsymbol{0}} \sup_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \boldsymbol{c}^\top \boldsymbol{x} - \sum_{i\in[I]} y_i f_i\left(\boldsymbol{a^i}, \boldsymbol{x}\right) \right\} = \inf_{\boldsymbol{y}\geq\boldsymbol{0}} \left(\sum_{i\in[I]} y_i f_i \right)^* \left(\boldsymbol{a^i}, \boldsymbol{c}\right)$$
$$= \inf_{\boldsymbol{y}\geq \boldsymbol{0}, \boldsymbol{w^i}\in\mathbb{R}^n} \left\{ \sum_{i\in[I]} y_i f_i^* \left(\boldsymbol{a^i}, \frac{\boldsymbol{w^i}}{y_i} \right) : \sum_{i\in[I]} \boldsymbol{w^i} = \boldsymbol{c} \right\}, \quad (9)$$

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:

$$\text{(OD)} \quad \inf_{\boldsymbol{y}\geq\boldsymbol{0},\, \boldsymbol{a^i}\in\mathbb{R}^L,\, \boldsymbol{w^i}\in\mathbb{R}^n} \sum_{i\in[I]} y_i f_i^* \left(\boldsymbol{a^i}, \frac{\boldsymbol{w^i}}{y_i} \right)$$
$$\text{s.t.} \quad \sum_{i\in[I]} \boldsymbol{w^i} = \boldsymbol{c}$$
$$g_{ik}\left(\boldsymbol{a^i}\right) \leq 0 \qquad \forall i \in [I] \quad \forall k \in [K_i]. \quad (10)$$

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 $y_i f_i^*(\boldsymbol{a^i}, \boldsymbol{w^i}/y_i)$. However, this problem has exactly the hidden convexity structure studied in this paper. An equivalent but convex problem (COD) can therefore be obtained by substituting $y_i \boldsymbol{a^i} = \boldsymbol{v^i}$ and multiplying constraint (10) with y_i (cf. Gorissen et al. 2014):

$$\text{(COD)} \quad \inf_{\boldsymbol{y}\geq\boldsymbol{0}, \boldsymbol{v^i}\in\mathbb{R}^L, \boldsymbol{w^i}\in\mathbb{R}^n} \sum_{i\in[I]} y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i} \right)$$
$$\text{s.t.} \quad \sum_{i\in[I]} \boldsymbol{w^i} = \boldsymbol{c} \quad (11)$$
$$y_i g_{ik} \left(\frac{\boldsymbol{v^i}}{y_i} \right) \leq 0 \qquad \forall i \in [I] \quad \forall k \in [K_i]. \quad (12)$$

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, and it is proven along the same lines.

Theorem 2 *Assume (COD) satisfies the Slater condition and (COD) is bounded, then (COD)* *and (RC) have the same optimal value. Moreover, the part of the KKT vector of (COD) that corresponds to constraint* (11) *gives an optimal solution of (RC).*

Proof. Since (COD) is convex and Slater regular, it has a KKT vector $(\boldsymbol{x}, \boldsymbol{z})$ (Rockafellar 1970, Thm. 28.2). The KKT vector solves (Rockafellar 1970, Cor. 28.4.1):

$$\max_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\geq\boldsymbol{0}} \quad \inf_{\boldsymbol{y}\geq\boldsymbol{0}, \boldsymbol{v^i}\in\mathbb{R}^L, \boldsymbol{w^i}\in\mathbb{R}^n} \sum_{i\in[I]} y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i}\right) + \boldsymbol{x}^T \left(\boldsymbol{c} - \sum_{i\in[I]} \boldsymbol{w^i}\right) + \sum_{i\in[I]} \sum_{k\in[K_i]} z_{ik} y_i g_{ik} \left(\frac{\boldsymbol{v^i}}{y_i}\right).$$

Since the uncertainty region is bounded, the substitution $y_i \boldsymbol{a^i} = \boldsymbol{v^i}$ is reversible (Gorissen et al. 2014, Lemma 1) and the KKT vector indeed solves (RC):

$$\begin{aligned}
&\max_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\geq\boldsymbol{0}} \quad \inf_{\boldsymbol{y}\geq\boldsymbol{0}, \boldsymbol{a^i}\in\mathbb{R}^L, \boldsymbol{w^i}\in\mathbb{R}^n} \sum_{i\in[I]} y_i f_i^* \left(\boldsymbol{a^i}, \frac{\boldsymbol{w^i}}{y_i}\right) + \boldsymbol{x}^T \left(\boldsymbol{c} - \sum_{i\in[I]} \boldsymbol{w^i}\right) + \sum_{i\in[I]} \sum_{k\in[K_i]} z_{ik} g_{ik} \left(\boldsymbol{a^i}\right) \\
&= \max_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ \boldsymbol{c}^T\boldsymbol{x} + \inf_{\boldsymbol{y}\geq\boldsymbol{0}, \boldsymbol{a^i}\in\mathbb{R}^L} \left\{ -\sum_{i\in[I]} y_i f_i^{**} \left(\boldsymbol{a^i}, \boldsymbol{x}\right) : \boldsymbol{a^i} \in \mathcal{U}_i \right\} \right\} \\
&= \max_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{z}\geq\boldsymbol{0}} \left\{ \boldsymbol{c}^T\boldsymbol{x} : f_i\left(\boldsymbol{a^i}, \boldsymbol{x}\right) \leq 0 \quad \forall i \in [I] \ \ \forall \boldsymbol{a^i} \in \mathcal{U}_i \right\}.
\end{aligned}$$

■

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 C_i be a closed convex cone, and suppose the uncertainty region $\mathcal{U}_i$ is described by one or more conic inclusion constraints: $\boldsymbol{D^i a^i} - \boldsymbol{d^i} \in C_i$ 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* (12) *is $\boldsymbol{D^i v^i} - y_i \boldsymbol{d^i} \in C_i$.*

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 $y_i > 0$, $y_i g_{ik}(\boldsymbol{v^i}/y_i) \leq 0$ if and only if $g_{ik}(\boldsymbol{v^i}/y_i) = 0$, i.e., $\boldsymbol{D^i v^i}/y_i - \boldsymbol{d^i} \in C_i$. Since C_i is a cone, this is equivalent to $\boldsymbol{D^i v^i} - y_i \boldsymbol{d^i} \in C_i$. When $y_i = 0$, $y_i g_{ik}(\boldsymbol{v^i}/y_i) \leq 0$ 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 $(0, y_i^*]$, $\boldsymbol{D^i v^i} - \varepsilon \boldsymbol{d^i} \in C_i$. Since C_i is closed, $\boldsymbol{D^i v^i} - y_i \boldsymbol{d^i} \in C_i$.

■

Our method requires tractable convex conjugates of the constraint functions. In Appendix A, 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. Surprisingly, as shown in Appendix A, the conjugates are often conic representable even if the uncertain parameters get added as optimization variables.

3.3 Advantages of the convex dual formulation

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

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 (7) as a maximization problem over $\boldsymbol{a^i}$ as shown in (8). Assuming $\mathrm{ri}(\mathrm{dom}(\mathcal{U}_i)) \cap \mathrm{ri}(\mathrm{dom}(f_i(\cdot, \boldsymbol{x}))) \neq \emptyset$, dualizing each constraint of (RC) using Fenchel's duality theorem yields the following equivalent convex optimization problem (Ben-Tal et al. 2015):

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

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., the concave function $f(a) = \exp(-a^2)$ if $a \in [0, 1/\sqrt{2}]$, $f(a) = -\infty$ otherwise. Many (convex) uncertainty regions therefore previously led to intractable RCs for 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

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

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. 2015), 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.

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. We demonstrate this in Section A.1 with the example of a robust geometric optimization problem. The primal problem is either badly scaled or has a dense Hessian, while the dual is well scaled and has a diagonal Hessian.

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 algorithms. 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 from an optimal solution.

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.

4 Inverse Optimization

Inverse optimization is about recovering optimization parameters from an optimal solution. Often the goal is to find the coefficients in the objective function, but in this section we also consider unknown coefficients in the constraints. Using the hidden convexity reformulation technique introduced in Section 2, we present inverse optimization models for linear and nonlinear optimization problems.

4.1 Linear optimization problems

Given an optimal solution $\hat{\boldsymbol{x}}$ to the general linear optimization problem

$$\begin{aligned} \sup_{\boldsymbol{x}\geq\boldsymbol{0}} \quad & \boldsymbol{c}^\top\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x}\leq\boldsymbol{b}, \end{aligned}$$

with $\boldsymbol{c}\in\mathbb{R}^n$, $\boldsymbol{b}\in\mathbb{R}^m$ and $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, we recover (parts of) $\boldsymbol{A}$ and $\boldsymbol{c}$ by using an alternative version of the duality gap model in Chan and Kaw (2020). Their approach relies on dual feasibility constraints that hold with equality ($\boldsymbol{A^T y}=\boldsymbol{c}$), so $\boldsymbol{c}$ can be eliminated from the problem. This approach does not work when (i) $\boldsymbol{c}$ is known or (ii) some information on $\boldsymbol{c}$ is known (i.e., when there are side constraints on $\boldsymbol{c}$). We propose an alternative approach that resolves these issues by using the hidden convexity structure.

A first difference with respect to Chan and Kaw (2020) is that we consider some possible convex constraints on $\boldsymbol{c}$, i.e., $\boldsymbol{c}\in\Gamma$. Secondly, we assume the convex constraints on $\boldsymbol{A}$ to be of a columnwise structure: $g_{ik}(\boldsymbol{a^i})\leq 0,\ i\in[m],\ k\in[K_i]$, where $g_{ik}(\cdot)$ is separable in $\boldsymbol{a^1},\boldsymbol{a^2},\ldots,\boldsymbol{a^m}$. We arrive at the nonconvex inverse optimization model, that enforces primal and dual feasibility and minimizes the duality gap:

$$\min_{\boldsymbol{a^i},\boldsymbol{y}\geq\boldsymbol{0},\boldsymbol{c}} \left\{ \boldsymbol{b}^T\boldsymbol{y} - \boldsymbol{c}^T\hat{\boldsymbol{x}} \ : \ \boldsymbol{c}\in\Gamma,\ \sum_{i\in[m]} y_i\boldsymbol{a^i} \geq \boldsymbol{c},\ (\boldsymbol{a^i})^T\hat{\boldsymbol{x}} \leq b_i,\ g_{ik}(\boldsymbol{a^i}) \leq 0,\ \forall i\ \forall k \right\},$$

where $\boldsymbol{y}$ refers to the dual vector corresponding to the constraints of the primal problem. We recognize that this problem has the hidden convexity structure studied in this paper. Using the substitution $\boldsymbol{v^i} = y_i\boldsymbol{a^i}$, we arrive at the convex reformulation

$$\min_{\boldsymbol{v^i},\boldsymbol{y}\geq\boldsymbol{0},\boldsymbol{c}} \left\{ \boldsymbol{b}^T\boldsymbol{y} - \boldsymbol{c}^T\hat{\boldsymbol{x}} : \boldsymbol{c}\in\Gamma,\ \sum_{i\in[m]} \boldsymbol{v^i} \geq \boldsymbol{c},\ (\boldsymbol{v^i})^T\hat{\boldsymbol{x}} \leq b_iy_i,\ y_ig_{ik}(\boldsymbol{v^i}/y_i) \leq 0,\ \forall i\ \forall k \right\}. \quad (13)$$

Compared to Chan and Kaw (2020), the inverse optimization model (13) holds several advantages. First, it can also be used when $\boldsymbol{c}$ is known or when information on $\boldsymbol{c}$ is known. Additionally, our method is extendable to nonlinear optimization problems as presented in Section 4.2 and does not require the normalization constraint as in Chan and Kaw (2020).

4.2 Nonlinear optimization problems

We extend the inverse optimization approach for *linear* problems of the previous section to a broad class of *nonlinear* optimization problems. We consider the following class of nonlinear optimization problems:

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

where $\boldsymbol{a^i} \in \mathcal{U}_i \in \mathbb{R}^L$, $\forall i \in I$, is an unknown parameter and $f_i : \mathbb{R}^L \times \mathbb{R}^n \to \mathbb{R}$ are the constraint functions. We assume that I is finite, and for each i, that $f_i(\boldsymbol{a^i},\boldsymbol{x})$ is linear 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.

Before explaining the inverse optimization approach, we first describe several examples of optimization problems that fit into the structure of (NP). The first example are convex quadratic constraints.

Example 3 *Consider* $f_i(\boldsymbol{a^i}, \boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A^i} \boldsymbol{x} + (\boldsymbol{b^i})^T \boldsymbol{x} + c_i$*, in which the parameters* $\boldsymbol{A^i} \in \mathbb{S}^n_+$*,* $\boldsymbol{b^i}$*, and* c_i *might be unknown. This function is in the format of problem (NP) since it is linear in the unknown parameters, and convex in the optimization variable.*

The next example is to learn the weights in the weighted-sum approach for multi-objective optimization.

Example 4 *One of the approaches in multi-objective optimization is the weighted-sum approach. In this approach the final objective function is the weighted sum of all the objectives. This approach is, for example, used in many hospitals to find optimal radiotherapy treatment plans for cancer patients. Inverse optimization can be used to find weights that have been implicitly or explicitly used in the past. In case the multiple objectives are nonlinear convex functions* $g_i(\boldsymbol{x})$*, the final weighted-sum objective is* $\sum_i w_i g_i(\boldsymbol{x})$*. This function, where* $\boldsymbol{w}$ *has to be learned, satisfies the format of problem* (NP)*.*

Suppose that $\hat{\boldsymbol{x}}$ is a known optimal solution to problem (NP). Since f_i, $i \in I$, is closed proper convex, we may use the dual formulation (9). Using the hidden convexity substitution, $\boldsymbol{v^i} = y_i \boldsymbol{a^i}$, we arrive at the convex inverse optimization model, that enforces primal and dual feasibility, and minimizes the duality gap:

$$\min_{\boldsymbol{v^i}, \boldsymbol{w^i}, \boldsymbol{y} \geq \boldsymbol{0}, \boldsymbol{c}} \left\{ \sum_{i \in I} y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i} \right) - \boldsymbol{c}^T \hat{\boldsymbol{x}} \ : \ \boldsymbol{c} \in \Gamma, \ \sum_{i \in I} \boldsymbol{w^i} = \boldsymbol{c}, \right.$$
$$\left. y_i f_i \left(\frac{\boldsymbol{v^i}}{y_i}, \hat{\boldsymbol{x}} \right) \leq 0, \ \ y_i g_{ik} \left(\frac{\boldsymbol{v^i}}{y_i} \right) \leq 0, \ \ \forall i \, \forall k \right\}.$$

where $\boldsymbol{y}$ refers to the dual vector corresponding to the constraints of the primal problem. We observe that both $y_i f_i^* \left(\frac{\boldsymbol{v^i}}{y_i}, \frac{\boldsymbol{w^i}}{y_i} \right)$ and $y_i f_i \left(\frac{\boldsymbol{v^i}}{y_i}, \hat{\boldsymbol{x}} \right)$ are convex in $(\boldsymbol{v^i}, \boldsymbol{w^i}, y_i)$ since $f_i(\boldsymbol{a^i}, \boldsymbol{x})$ is convex in $\boldsymbol{x}$ and linear in $\boldsymbol{a^i}$.

5 Numerical Example: Food Supply Chain For World Food Programme

In this section we consider a case study on the United Nations World Food Programme (WFP) to illustrate three applications belonging to the class of optimization problems discussed in this paper. The case concerns an optimization model for the food assistance response by the

WFP which includes the food basket to be delivered, the sourcing plan and the delivery plan of a recovery operation. The food basket should be composed of an optimal collection of *commodities* (i.e., raw goods or primary agricultural products), where minimum requirements are enforced on the contained *nutrients* (i.e., chemical compounds in food items that are used by the body to maintain health). Peters et al. (2021) have modeled such a mixed integer linear programming model which we adopt in this paper, and their model is therefore explained in Section 5.1. We treat three extensions to this model that lead to nonconvex problems that have the hidden convexity structure studied in this paper. Hence, we can use the reformulation technique proposed in this paper to reformulate the problems into convex ones. These extensions concern variable coefficients, a best-case scenario and inverse optimization, which are described in Sections 5.2, 5.3 and 5.4, respectively. Along a presentation of the theoretical extensions, these sections include numerical results using data from the civil war in Syria in 2017.

5.1 The WFP model

The WFP offers food assistance to over 115 million people each year, reaching out to people in humanitarian crisis and hunger. In collaboration with the WFP, Peters et al. (2021) have developed a model that combines dietary optimization with supply chain optimization. After receiving successful test results in Syria, it is now implemented in WFP's internal systems and used worldwide. The problem consists of collecting food from suppliers in the food industry and delivering it to food relief agencies that serve individuals in need, of which an example is illustrated in Figure 1. While previously transportation decisions at WFP relied on a predetermined food basket, Peters et al. (2021) propose a variable food basket. That is, optimizing the product compositions per beneficiary type, rather than enforcing fixed mixtures of foods. The intention of the model is to serve all beneficiaries in need and to select the right ingredients for their daily ration, such that costs are minimal and the nutritional requirements are met.

A simplified version of the mixed integer linear programming model by Peters et al. (2021) that

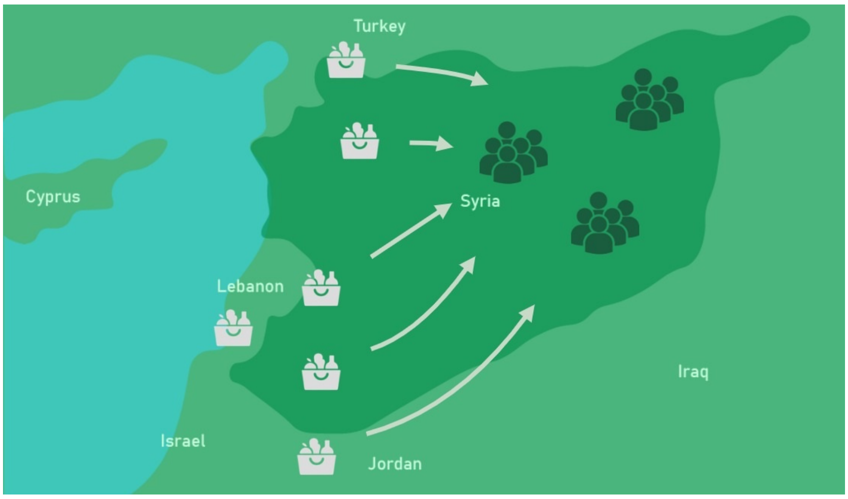

Figure 1: An example of the food collection and distribution problem in Syria, to serve the hungry population in Ar Raqqah, Al-Qamishli and Deir ez-Zor. There are regional suppliers as well as international suppliers from Lebanon, Jordan and Turkey.

we use is:

$$
\begin{aligned}
\text{(M)} \quad \min_{F_{ijk},R_k} \quad & \sum_{i \in i \in \mathcal{N}_\mathcal{S}} \sum_{j \in \mathcal{N}_\mathcal{D}} \sum_{k \in \mathcal{N}_\mathcal{C}} \left(p^P_{ik} + p^T_{ijk}\right) F_{ijk} \\
\text{s.t.} \quad & \sum_{i \in \mathcal{N}_\mathcal{S}} F_{ijk} = d_j R_k && \forall j \in \mathcal{N}_\mathcal{D} \quad \forall k \in \mathcal{N}_\mathcal{C} && (14) \\
& \sum_{k \in \mathcal{N}_\mathcal{C}} nutval_{kl} R_k \geq nutreq_l && \forall l \in \mathcal{N}_\mathcal{N} && (15) \\
& F_{ijk}, R_k \geq 0 && \forall i \in \mathcal{N}_\mathcal{S} \quad \forall j \in \mathcal{N}_\mathcal{D} \quad \forall k \in \mathcal{N}_\mathcal{C}. && (16)
\end{aligned}
$$

The notation is the same as in Peters et al. (2021):

Type	Symbol	Description
Set	$\mathcal{N}_\mathcal{S}$	Suppliers
Set	$\mathcal{N}_\mathcal{D}$	Demand locations
Set	$\mathcal{N}_\mathcal{C}$	Commodities
Set	$\mathcal{N}_\mathcal{N}$	Nutrients
Parameter	d_j	Number of beneficiaries at location $j \in \mathcal{N}_\mathcal{D}$
Parameter	$nutreq_l$	Nutritional requirement per person per day for nutrient $l \in \mathcal{N}_\mathcal{N}$
Parameter	$nutval_{kl}$	Amount of nutritional value in commodity $k \in \mathcal{N}_\mathcal{C}$ of nutrient $l \in \mathcal{N}_\mathcal{N}$
Parameter	p^P_{ik}	Procurement costs (in \$/100 g) at location $i \in \mathcal{N}_\mathcal{S}$ for commodity $k \in \mathcal{N}_\mathcal{C}$
Parameter	p^T_{ijk}	Costs of transporting (in \$/100 g) from location $i \in \mathcal{N}_\mathcal{S}$ to location $j \in \mathcal{N}_\mathcal{D}$ of commodity $k \in \mathcal{N}_\mathcal{C}$
Variable	F_{ijk}	Amount of commodity $k \in \mathcal{N}_\mathcal{C}$ shipped from location $i \in \mathcal{N}_\mathcal{S}$ to location $j \in \mathcal{N}_\mathcal{D}$ in 100 grams
Variable	R_k	Ration of commodity $k \in \mathcal{N}_\mathcal{C}$ in the daily food basket in 100 grams.

The objective is to minimize costs, consisting of procuring and transporting costs. Balance constraints (14) equalize the arriving commodity flow at the beneficiaries to the demanded rations per commodity. Constraint (15) provides that the supplied nutrients meet a minimal requirement and constraint (16) ensures the decision variables are positive. Model (M) finds the optimal solution for a single planning period.

5.2 Application 1: variable nutritional values

The parameter $nutval_{kl}$ in constraint (15) specifies the amount of nutrient l in commodity k. There exist, however, products whose nutritional value is not fixed, meaning that the same product can be ordered with different nutritional values for the same price. An example of such a product is milk powder, where the amount of each ingredient can be tuned. We propose to incorporate this in the model and take the nutritional components of such commodities as variables. This adjustment allows the WFP to align the ingredients of such commodities to the nutritional needs of the beneficiary location. Using this extension, it becomes possible to order, for instance, protein-rich milk powder for regions struggling to meet the protein requirements.

For this application we consider the nutritional values of a predefined set of commodities as variable. We define $\mathcal{N}_{\mathcal{V}} \subset \mathcal{N}_{\mathcal{C}}$ as the collection of commodities with variable nutritional value, $\mathcal{C}_k \subset \mathbb{R}^{|\mathcal{N}_{\mathcal{N}}|}$ as the convex set of permissible nutritional values, and V_{kl} as the variable indicating the amount of nutrient l in commodity k. Then (15) is modified into the following constraints:

$$\sum_{k \in \mathcal{N}_{\mathcal{C}} \setminus \mathcal{N}_{\mathcal{V}}} nutval_{kl} R_k + \sum_{k \in \mathcal{N}_{\mathcal{V}}} V_{kl} R_k \geq nutreq_l \qquad \forall l \in \mathcal{N}_{\mathcal{N}} \tag{17}$$

$$\boldsymbol{V^k} \in \mathcal{C}_k \qquad \forall k \in \mathcal{N}_{\mathcal{V}}. \tag{18}$$

Constraint (17) contains the bilinear term $V_{kl}R_k$ and demonstrate the hidden convexity structure as discussed in this paper.

This model extension is tested on real data from the civil war in Syria in 2017. The data originates from Poos (2020), who has composed the data in collaboration with the WFP. We consider three international and five regional suppliers to give food assistance to people in hunger stationed at five locations in Syria. A collection of these locations is depicted in Figure 1. By composing a food basket choosing from 25 commodities, we will meet requirements for 11 nutrients. A planning period of one day is considered. We refer to this setting as the nominal model, for which the results can be found in Table 2.

For the first model extension, we consider 'dried skim milk' as the only commodity in the set $\mathcal{N}_{\mathcal{V}}$. That is, the nutritional components of dried skim milk are taken variable, adding an additional 11 variables to the nominal model ($V_{kl},\ k \in \mathcal{N}_{\mathcal{V}}$). The reformulation technique is used for the nonconvex constraints (17). For $\mathcal{C}_k$ we consider polyhedral regions:[1]

$$(1-\theta) nutval_{kl} \leq V_{kl} \leq (1+\theta) nutval_{kl} \qquad \forall k \in \mathcal{N}_{\mathcal{C}} \quad \forall l \in \mathcal{N}_{\mathcal{N}} \tag{19}$$

$$\sum_{l \,\in\, \mathcal{N}_{\mathcal{L}_k}} \rho \, \frac{V_{kl}}{nutval_{kl}} \leq |\mathcal{N}_{\mathcal{L}_k}| \qquad \forall k \in \mathcal{N}_{\mathcal{C}}, \tag{20}$$

where $\mathcal{N}_{\mathcal{L}_k}$ indicate the sets of nutrients per commodity k for which $nutval_{kl} \neq 0$, θ is the permissible relative deviation from the nominal value $nutval_{kl}$ and ρ the scalar for the linear relation among the nutritional components of a given commodity. When the value of ρ is equal to 1 the half planes in (20) cut through the nominal point $nutval_{kl}$. This ensures that there are no combinations allowed that are unrealistically more nutritious on all levels. For this application we take $\theta = 0.2$ and $\rho = 1$. The results of this first extension can be found in Table 2, which shows that the adjusted food basket composition results in a 6.2% reduction in costs, while still meeting the same nutritional requirements, compared to the nominal case. The gain is made by aligning the nutritional components of the dried skim milk to the beneficiary nutritional needs. That this is a good strategy is also visible from the food basket composition: dried skim milk, the only commodity with variable nutrients, becomes a more important element of the basket.

Table 2: Summary of results of the nominal model and extended model with variable nutritional values. The cost of the optimal delivery plan and the nonzero rations are included in this table.

	Nominal	Variable coefficients
Total cost (USD)	25,571	23,991
Cost per person (USD)	0.332	0.312
Ration values R_k (g)		
Corn-soya blend	171	0
Dried skim milk	18	213
Soya-fortified bulgur wheat	204	43
Oil	76	81
Wheat-soya blend	0	84

[1] Alongside a polyhedral region for $\mathcal{C}_k$, we have also tested for a box-ellipsoidal region with the origin as the center for the ellipse. Hence, (19) combined with $\sum_{l \,\in\, \mathcal{N}_{\mathcal{L}_k}} (\rho\, V_{kl}/nutval_{kl})^2 \leq |\mathcal{N}_{\mathcal{L}_k}|,\ \forall\, k \in \mathcal{N}_{\mathcal{C}}$. However, the polyhedral and box-ellipsoidal region resulted in the same outcomes for $\theta = 0.2$ and $\rho = 1$, which motivated the decision to solely state the polyhedral region in the body of this paper.

5.3 Application 2: best-case solution

The WFP model in (M) assumes that the precise nutritional values of all commodities are known. In practice, this is hardly the case as the nutritional value is affected by: the origin of the product (e.g., the soil quality), the preparation method (e.g., baking, boiling or raw consumption) and the digestive system of the beneficiary. Hence, we deal with intervals of the nutritional profile rather than single values and so there is uncertainty regarding the nutrient values. Though the size of the uncertainty range strongly depends on the type of commodity, the *exact* nutritional profiles always contain a degree of uncertainty. As a second extension of the model, we propose a method that contributes information on the possible solutions while considering such uncertainty ranges. In contrast to Robust Optimization, which yields a solution that is optimal for the worst-case scenario, we look at the optimal solution for the best-case scenario. Particularly, we modify model (M) such that it will find an optimal solution for the model when the nutritional values of all commodities are equal to the most ideal value. Hence, it presents a lower bound on the objective which can be useful information for the decision maker. When combined with a robust optimum, it gives a range on the possible variations in the objective.

For this application we consider the coefficient $nutval_{kl}$ as variable, $\forall k \in \mathcal{N}_\mathcal{V}$ $\forall l \in \mathcal{N}_\mathcal{N}$. We adopt the notation and constraint modifications as for Application 1, while extending the variability to *all* commodities k. Hence, (15) is replaced by constraints (17)–(18) with $\mathcal{N}_\mathcal{V} = \mathcal{N}_\mathcal{C}$. These modified constraints contain bilinear terms $V_{kl}R_k$ and demonstrate the hidden convexity structure as discussed in this paper.

This second application is also executed on the Syria case, for which we add $|\mathcal{N}_\mathcal{C}| \times |\mathcal{N}_\mathcal{N}|$ new variables to the nominal model and use the presented reformulation on nonconvex terms $V_{kl}R_k$. For $\mathcal{C}_k$ we use polyhedral regions as specified in (19)–(20), with $\theta = 0.2$. We take $\rho = 1.1$, allowing values of nutrients to be chosen that are, in absolute terms, greater than the nominal value. We also compute the worst-case solution using Robust Optimization with the set $\mathcal{C}_k$ as the uncertainty set.

We find that the food basket of the best-case solution is composed of the exact same commodities as in the nominal case, but with different quantities. Presumably, the freedom of selecting in the nutritional profiles of the other commodities is of minor importance. Results of the best-case application are compared to the nominal solution and a robust solution in Table 3. The table presents the total costs as well as the solutions' feasibility performance for the nominal scenario, and the best and worst case scenario for that specific solution. The latter is measured using nutritional value scores (NVSs) in order to quantify the degree of violation of the nutritional constraints (17). The NVS is equal to the sum of delivered percentages (truncated at 100%) for each nutrient l:

$$\frac{1}{|\mathcal{N}_\mathcal{N}|}\sum_{l\in\mathcal{N}_\mathcal{N}} \min\left\{100\%, \frac{\sum_{k\,\in\,\mathcal{N}_\mathcal{C}} V_{kl}R_k}{nutreq_l}\right\}.$$

The results show that the worst-case solution is always feasible, but this security results naturally in the highest total cost. With the nominal solution there is a chance that you will receive only 91% of the nutrients and is therefore a little bit more risky, while offering a better price. The highest risk and lowest total cost are associated with the best-case solution. This solution also yields the absolute lower bound on the objective.

Table 3: Summary of results for the best-case model extension. Total costs are compared to the nominal model and worst-case solution. The nutritional value score (NVS) is used as an indication for the solutions' performance per scenario. Note that the total costs do not depend on the realized nutritional values.

	Solution		
	Nominal	Best case	Worst case
Total cost (USD)	25,571	21,664	31,964
Nutritional value score			
Nominal	100%	94%	100%
Best case	100%	100%	100%
Worst case	91%	83%	100%

5.4 Application 3: approximate palatability constraints

A number of food baskets suggested by model (M), though meeting nutritional requirements, may consist of unpalatable mixtures of foods. A plate filled with solely a very large portion of rice, milk powder and oil, for instance, is not an adequate meal composition. It can be challenging to capture rules that exclude such unpalatable mixtures into constraints because of two reasons: the rules are (1) not explicitly known, (2) of high complexity. As a final application we propose a method of approximating palatability constraints with linear constraints. We consider a setup in which we know that one or more previous solutions contain a palatable-wise satisfactory food basket, while the exact formulation of the palatability constraints may be unknown. By extracting information from the existing solution(s) using inverse optimization, we create linear approximations of palatability constraints. This approach offers WFP the opportunity to learn constraints that preclude food combinations that are not palatable and disallow them in future supply chain decisions.

For this final application we define $\mathcal{N}_{\mathcal{T}}$ as the set of palatability constraints and assume that these constraints are linear in R_k. We refer to P_{kb} as the unknown palatability coefficient for commodity k in palatability constraint $b \in \mathcal{N}_{\mathcal{T}}$, and $\boldsymbol{P^b}$ as the vector of coefficients. Convex constraints on $\boldsymbol{P^b}$ are expressed using region $\mathcal{C}_b$. We extend model (M) with the following constraints:

$$\sum_{k \in \mathcal{N}_{\mathcal{C}}} P_{kb} R_k \geq 1 \qquad \forall b \in \mathcal{N}_{\mathcal{T}} \tag{21}$$

$$\boldsymbol{P^b} \in \mathcal{C}_b \qquad \forall b \in \mathcal{N}_{\mathcal{T}}.$$

Without loss of generality the right-hand-side of constraint (21) is normalized to 1. To numerically validate the inverse optimization approach, we take the nominal solution for R_k and assume that the corresponding food basket composition is palatable, and use the inverse optimization method from Section 4.1, to recover the implicit values of the coefficients P_{kb}. Note that for this application the objective coefficients in (M) remain fixed, and hence the inverse optimization method of Chan and Kaw (2020) cannot be used.

This third application is executed on the Syria case by using the WFP sensible food basket information (Peters et al. 2021, Table A.1) as a test environment for the palatability constraints. These sensibility guidelines provide lower bounds on the amount of Cereals & Grains (CG) and Pulses & Vegetables (PV) in each basket. Using the set notation $\mathcal{N}_{\mathcal{CG}}$ and $\mathcal{N}_{\mathcal{PV}}$ for the commodities belonging to each of these food groups, the specific sensible palatability constraints are:

$$\sum_{k \in \mathcal{N}_{\mathcal{CG}}} R_k \geq 2.5 \text{ and } \sum_{k \in \mathcal{N}_{\mathcal{PV}}} R_k \geq 0.3.$$

Since the right-hand-side of (21) is 1, the coefficients to be recovered are $1/2.5 = 0.4$ for the first constraint and $1/0.3 \approx 3.33$ for the second constraint (Table 4).

For $\mathcal{C}_b$ we consider a polyhedral region similar to (19)–(20), with $\theta = 0.2$, $\rho = 1$ and P_{kb} nonnegative. The exact expression for $\mathcal{C}_b$ is included in Appendix C.

We perform three tests: (1) solely retrieving a lower bound on Cereals & Grains (CG), (2) solely retrieving a lower bound on Pulses & Vegetables (PV), and (3) retrieving the lower bounds on CG and PV simultaneously. Since this is just a computational experiment, we have the ground trutht and can evaluate the results based on the mean absolute error (MAE):

$$\frac{\sum_{k\in\mathcal{N}_\mathcal{C}} |P_{kb} - \hat{P}_{kb}|}{|\mathcal{N}_\mathcal{C}|},$$

and on the weighted average percentage error (WAPE):

$$\frac{\sum_{k\in\mathcal{N}_\mathcal{C}} |P_{kb} - \hat{P}_{kb}|}{\sum_{k\in\mathcal{N}_\mathcal{C}} P_{kb}}.$$

Table 4 shows the approximated palatability parameters ($\hat{P}_{kb}$), the actual values (P_{kb}), and their corresponding MAE and WAPE. These results show that the proposed method performs relatively well on approximating the constraint coefficients. In particular when we aim to retrieve the coefficients for the CG constraint, the accuracy measure displays tolerable deviation levels. The performance of this method is not harmed when approximating multiple constraints simultaneously, which can be stated by comparing test 3 to tests 1 and 2. In conclusion, this application shows that it is possible for WFP to approximate palatability constraints by linear constraints using the proposed method.

Acknowledgment

The authors thank F. J. C. T. de Ruiter (Tilburg University) for his comments on a previous version of this paper and the company Hoogwegt (Arnhem, Netherlands), milk supplier for the WFP, for providing practical information on the first case study.

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.

R. K. Ahuja and J. B. Orlin. Inverse optimization. *Operations Research*, 49(5):771–783, 2001.

Table 4: Summary of results of approximating palatability constraints. The actual values (P_{kb}) are presented along with the inverse optimization results ($\hat{P}_{kb}$) on three test cases: (1) solely retrieving a lower bound on Cereals & Grains (CG), (2) solely retrieving a lower bound on Pulses & Vegetables (PV), and (3) retrieving the lower bounds on CG and PV simultaneously. To indicate the absolute and relative accuracy performance, the mean absolute error and weighted average percentage error are shown.

	Actual values		Test 1	Test 2	Test 3	
	CG	PV	CG	PV	CG	PV
Beans	0	3.33	0.03	3.13	0	2.67
Bulgur	0.40	0	0.34	0.31	0.32	0
Cheese	0	0	0	0	0	0
Fish	0	0	0.02	0.23	0	0
Meat	0	0	0	0	0	0
Corn-soya blend	0	0	0.07	0.08	0.04	0.05
Dates	0	3.33	0	3.18	0	2.67
Dried skim milk	0	0	0.02	0.03	0.05	0.04
Milk	0	0	0.04	0.37	0	0
Salt	0	0	0.04	0.36	0.04	0.46
Lentils	0	3.33	0.02	2.94	0.04	2.96
Maize	0.40	0	0.33	0.21	0.32	0
Maize meal	0.40	0	0.34	0.19	0.32	0
Chickpeas	0	3.33	0.03	2.82	0.05	2.71
Rice	0.40	0	0.34	0	0.32	0
Sorghum/millet	0.40	0	0.33	0.29	0.32	0
Soya-fortified bulgur wheat	0.40	0	0.38	0	0.34	0.02
Soya-fortified maize meal	0.40	0	0.32	0.10	0.43	0.05
Soya-fortified sorghum grits	0.40	0	0.40	0.06	0.34	0.09
Soya-fortified wheat flour	0.40	0	0.44	0.27	0.39	0.10
Sugar	0	0	0.03	0.21	0	0
Oil	0	0	0.01	0	0.02	0.04
Wheat	0.40	0	0.41	0.22	0.32	0
Wheat flour	0.40	0	0.41	0.35	0.32	0
Wheat-soya blend	0	0	0.04	0.12	0.02	0.05
MAE			0.03	0.19	0.04	0.13
WAPE			19%	35%	22%	24%

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

A. Ben-Tal and M. Teboulle. Hidden convexity in some nonconvex quadratically constrained quadratic programming. *Mathematical Programming*, 72:51–63, 1996.

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, USA, 2009.

A. Ben-Tal, D. den Hertog, and M. Laurent. Hidden convexity in partially separable optimization. *CentER Working Paper Series*, (70), 2011.

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 and D. den Hertog. *Robust and Adaptive Optimization*. Dynamic Ideas LLC, Belmont, MA, USA, 2022.

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. Reformulation versus cutting-planes for robust optimization. *Computational Management Science*, 13(2):195–217, 2015.

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

T. C. Chan and N. Kaw. Inverse optimization for the recovery of constraint parameters. *European Journal of Operational Research*, 282(2):415–427, 2020.

T. C. Chan, T. Lee, and D. Terekhov. Inverse optimization: Closed-form solutions, geometry, and goodness of fit. *Management Science*, 65(3):1115–1135, 2018.

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

I. Dunning. JuMPeR – robust optimization with JuMP, 2016. http://iainnz.github.io/JuMPeR.jl/, retrieved January 2021.

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.

G. Iyengar and W. Kang. Inverse conic programming with applications. *Operations Research Letters*, 33 (3):319–330, 2005.

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, NY, USA, 2009.

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

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

K. Peters, S. Silva, R. Gonçalves, M. Kavelj, H. Fleuren, D. den Hertog, O. Ergun, and M. Freeman. The nutritious supply chain: Optimizing humanitarian food assistance. *INFORMS Journal on Optimization*, 3(2):200–226, 2021.

R. Poos. A robust approach to food assistance supply chains. *Master Thesis Tilburg University*, 2020.

K. Postek, D. den Hertog, and B. Melenberg. Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review*, 58(4):603–650, 2016.

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

R. T. Rockafellar. Augmented lagrangians and hidden convexity in sufficient conditions for local optimality. *Mathematical Programming*, 1–36, 2022.

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

Z. Shahmoradi and T. Lee. Quantile inverse optimization: Improving stability in inverse linear programming. *Operations Research*, Published Online, 2021.

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.

Z. Y. Wu, D. Li, L. S. Zhang, and X. M. Yang. Peeling off a nonconvex cover of an actual convex problem: Hidden convexity. *SIAM Journal on Optimization*, 18(2):507–536, 2007.

J. Zhang and C. Xu. Inverse optimization for linearly constrained convex separable programming problems. *European Journal of Operational Research*, 200(3):671–679, 2010.

J. Zhang and L. Zhang. An augmented lagrangian method for a class of inverse quadratic programming problems. *Applied Mathematics and Optimization*, 61(1):57–83, 2010.

A Examples of conjugates

In this section, we derive conjugates for many types of constraint functions, 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.

A.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 $\boldsymbol{x} \geq \boldsymbol{0}$, $f(\boldsymbol{a}, \boldsymbol{x}) = \infty$ 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, 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 A.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:

$$\boldsymbol{\nu}^\top \boldsymbol{\Gamma}\boldsymbol{\nu} + \boldsymbol{\beta}^\top \boldsymbol{\nu} + \gamma \geq 0 \qquad \forall \boldsymbol{\nu} \in S. \tag{22}$$

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 (22) for any convex set S defined by finitely many convex functions.

Geometric optimization. A Geometric Optimization (GO) problem is characterized by posynomial constraints:

$$\sum_{i=1}^{k} c_i \prod_{j=1}^{n} x_j^{a_{ij}} \leq 1,$$

with $\boldsymbol{c} > \boldsymbol{0}$, $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ and $\boldsymbol{x} > \boldsymbol{0}$. Such a constraint can be equivalently expressed as:

$$\sum_{i=1}^{k} c_i \exp\left(\sum_{j=1}^{n} a_{ij} \log(x_j)\right) \leq 1,$$

which is convex in $\log(x_j)$, so substituting $x'_j = \log(x_j)$ yields a convex constraint. The substitution has detrimental effects on the scaling of the jacobian and hessian, so it is customary to apply a logarithmic transformation (Boyd and Vandenberghe 2004, §4.5.3):

$$\log\left(\sum_{i=1}^{k} c_i \exp\left((\boldsymbol{a^i})^\top \boldsymbol{x'}\right)\right) \leq 0. \tag{23}$$

Suppose there is rowwise uncertainty on $\boldsymbol{A}$, then the RC of (23) is:

$$\log\left(\sum_{i=1}^{k} c_i \exp(z_i)\right) \leq 0$$

$$(\boldsymbol{a^i})^\top \boldsymbol{x'} \leq z_i \qquad \forall \boldsymbol{a^i} \in \mathcal{U}_i \quad \forall i.$$

The disadvantage of this well-scaled constraint is that the hessian is dense. Deriving conjugate functions is straightforward:

$$f(\boldsymbol{z}) = \log\left(\sum_{i=1}^{k} c_i \exp(z_i)\right) \qquad f^*(\boldsymbol{w}) = \begin{cases} \sum_{i=1}^{k} w_i \log\left(\frac{w_i}{c_i}\right) & \text{if } \boldsymbol{w} \geq \boldsymbol{0}, \sum_{i=1}^{k} w_i = 1 \\ \infty & \text{otherwise} \end{cases}$$

$$g(\boldsymbol{a}, \boldsymbol{x}, z_j) = (\boldsymbol{a^i})^\top \boldsymbol{x} - z_i \qquad g^*(\boldsymbol{a}, \boldsymbol{u^i}, w_i) = \begin{cases} 0 & \text{if } \boldsymbol{u^i} = \boldsymbol{a^i}, w_i = -1 \\ \infty & \text{otherwise.} \end{cases}$$

On the effective domain of f, the first and second derivative of its perspective are:

$$\nabla y f^*\left(\frac{\boldsymbol{w}}{y}\right) = \begin{pmatrix} \log\left(\frac{w_1}{yc_1}\right) + 1 \\ \vdots \\ \log\left(\frac{w_k}{yc_k}\right) + 1 \\ -\sum_{i=1}^{k} \frac{w_i}{y} \end{pmatrix} \text{ and } \nabla^2 y f^*\left(\frac{\boldsymbol{w}}{y}\right) = \begin{pmatrix} \frac{1}{w_1} & & & \\ & \ddots & & \\ & & \frac{1}{w_k} & \\ & & & \frac{1}{y^2}\sum_{i=1}^{k} w_i \end{pmatrix}.$$

These functions are are well scaled, and the Hessian is sparse. Therefore, solving the dual is advantageous from a computational perspective.

A.2 Quadratic in the optimization variables

Consider the following two constraints:

$$\boldsymbol{x}^\top \boldsymbol{P}(\boldsymbol{\zeta})^\top \boldsymbol{P}(\boldsymbol{\zeta})\boldsymbol{x} + \boldsymbol{b}(\boldsymbol{\zeta})^\top \boldsymbol{x} + c(\boldsymbol{\zeta}) \leq 0 \qquad \forall \boldsymbol{\zeta} \in \mathcal{Z} \tag{24}$$

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} + c \leq 0 \qquad \forall (\boldsymbol{A}, \boldsymbol{b}, c) \in \mathcal{U}. \tag{25}$$

Ben-Tal et al. (2009, Chapter 6) focus on constraint (24) 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 largest singular value is 1.

We consider constraint (25), 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 (25) 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 $\boldsymbol{A^j}$, $\boldsymbol{b^j}$ 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 $\boldsymbol{V}$ and $\boldsymbol{F}^{-1}$. 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):

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

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: $yf^*(\boldsymbol{A}/y, \boldsymbol{b}/y, c/y, \boldsymbol{u}/y) \leq t$ is equivalent to:

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

Indeed y does not appear in the SDr. A more tractable result can be obtained when $\boldsymbol{A}$ is affinely parameterized by an uncertain nonnegative vector $\boldsymbol{s} \in S \subset \mathbb{R}^k_+$:

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

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. (2016) showed that the variance of $\boldsymbol{x}$ can be written as $\min_{\pi \in \mathbb{R}} \sum_{j=1}^{n} a_j (x_j - \pi)^2$, 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.

A.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:

$$p_t(\boldsymbol{x}) = t + h_j(\boldsymbol{x}) - (\boldsymbol{u^j})^\top \boldsymbol{x} \geq 0 \qquad \forall \boldsymbol{x} \in \mathbb{R}^n. \tag{26}$$

Constraint (26) 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 (26) 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 (26) 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.

A.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 $\left|\left| \boldsymbol{A}^{\frac{1}{2}} (\boldsymbol{x} - \boldsymbol{b}) \right|\right|_2 \leq \boldsymbol{c}^\top \boldsymbol{x} + d$, 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:

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

The constraint $\left|\left| \boldsymbol{A}^{-\frac{1}{2}} (\boldsymbol{u} + \boldsymbol{c}) \right|\right|_2 \leq 1$ 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:

$$f^*(\boldsymbol{a}, c, \boldsymbol{U}) = \inf_{\boldsymbol{Z} \in \mathbb{R}^{n\times n}} \{\delta^*(\boldsymbol{Z}|\mathbb{S}^n_+) - g_*(\boldsymbol{a}, c, \boldsymbol{U}, \boldsymbol{Z})\}. \tag{27}$$

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 (27) 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.

A.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:

$$g(\boldsymbol{a}, \boldsymbol{x}) \leq \inf_{\boldsymbol{a}' \in \mathcal{U}'} \{h(\boldsymbol{a}, \boldsymbol{a}')\} \qquad \forall \boldsymbol{a} \in \mathcal{U}, \tag{28}$$

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 (28) 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 (28) 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.

B 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). We let $\mathbf{1}$ denote the vector of all 1's, with dimension implied by the context. As an illustrative example, we consider the following problem:

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

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 A.4, A.2, and A.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:

$$\begin{pmatrix} 4\boldsymbol{V^2} & \boldsymbol{w^2} - y_2\boldsymbol{b} \\ (\boldsymbol{w^2} - y_2\boldsymbol{b})^\top & cy_2 + t \end{pmatrix} \in \mathbb{S}^{n+1}_+,$$

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 $g_{11} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, 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 (12), $y_1 g_{11}(\boldsymbol{V^1}/y_1) \leq 0$, 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 (11), we get:

$$
\begin{aligned}
\text{(COD-Example)} \quad \inf \quad & \boldsymbol{b}^\top(\boldsymbol{w^1} + y_1\boldsymbol{c}) + y_1 d + t - y_3 \\
\text{s.t.} \quad & \begin{pmatrix} \boldsymbol{V^1} & \boldsymbol{w^1} + y_1\boldsymbol{c} \\ (\boldsymbol{w^1} + y_1\boldsymbol{c})^\top & y_1 \end{pmatrix} \in \mathbb{S}_+^{n+1} \\
& \begin{pmatrix} 4\boldsymbol{V^2} & \boldsymbol{w^2} - y_2\boldsymbol{b} \\ (\boldsymbol{w^2} - y_2\boldsymbol{b})^\top & cy_2 + t \end{pmatrix} \in \mathbb{S}_+^{n+1} \\
& \boldsymbol{w}_i^{\boldsymbol{3}} \leq -|y_3^{\frac{1}{3}} - (\boldsymbol{v}_i^{\boldsymbol{3}})^{\frac{1}{3}}|^3 \quad \forall i = 1, \ldots, n \\
& \sum_{i=1}^n \boldsymbol{w^i} = \boldsymbol{c} \qquad (29) \\
& \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{V}_{ij}^{\boldsymbol{1}} - \bar{\boldsymbol{A}}_{ij} y_1)^2 \leq \rho y_1^2 \\
& y_2 \boldsymbol{I} - \boldsymbol{V^2} \in \mathbb{S}_+^n \\
& \boldsymbol{v}_i^{\boldsymbol{3}} \leq y_3 \quad \forall i = 1, \ldots, n \\
& \left\| \boldsymbol{v^3} - \frac{1}{2} y_3 \boldsymbol{1} \right\|_2 \leq \rho y_3 \\
& \boldsymbol{y} \in \mathbb{R}_+^3,\ \boldsymbol{V^1} \in \mathbb{R}^{n\times n},\ \boldsymbol{V^2} \in \mathbb{S}_+^n,\ \boldsymbol{v^3} \in \mathbb{R}_+^n,\ \boldsymbol{w^i} \in \mathbb{R}^n.
\end{aligned}
$$

An optimal solution $\boldsymbol{x}$ to (Example) is now given by the dual variable with respect to constraint (29) in an optimal solution of (COD-Example).

C Polyhedral region palatability extension

For the numerical example in Section 5.4, we use the following polyhedral region for $\mathcal{C}_b$:

$$(1-\theta)\hat{p}_{kb} \leq P_{kb} \leq (1+\theta)\hat{p}_{kb} \qquad \forall k \in \mathcal{N}_{\mathcal{K}_b} \quad \forall b \in \mathcal{N}_{\mathcal{T}} \qquad (30)$$

$$-\theta \sum_{k\in\mathcal{N}_{\mathcal{K}_b}} \hat{p}_{kb} \leq P_{kb} \leq \theta \sum_{k\in\mathcal{N}_{\mathcal{K}_b}} \hat{p}_{kb} \qquad \forall k \in \mathcal{N}_{\mathcal{C}} \setminus \mathcal{N}_{\mathcal{K}_b} \quad \forall b \in \mathcal{N}_{\mathcal{T}} \qquad (31)$$

$$\sum_{k\in\mathcal{N}_{\mathcal{K}_b}} \rho \frac{P_{kb}}{\hat{p}_{kb}} \leq |\mathcal{N}_{\mathcal{K}_b}| \qquad \forall b \in \mathcal{N}_{\mathcal{T}} \qquad (32)$$

$$P_{kb} \geq 0 \qquad \forall k \in \mathcal{N}_{\mathcal{C}} \quad \forall b \in \mathcal{N}_{\mathcal{T}}, \qquad (33)$$

where $\hat{p}_{kb}$ indicates the nominal value for the palatability coefficient and $\mathcal{N}_{\mathcal{K}_b}$ the set of commodities for which $\hat{p}_{kb} \neq 0$.