

# Extending the Scope of Robust Quadratic Optimization

Ahmadreza Marandi<sup>\*†</sup>  
Aharon Ben-Tal<sup>‡</sup>  
Dick den Hertog<sup>§</sup>  
Bertrand Melenberg<sup>§</sup>

September 4, 2019

## Abstract

We derive computationally tractable formulations of the robust counterparts of convex quadratic and conic quadratic constraints that are concave in matrix-valued uncertain parameters. We do this for a broad range of uncertainty sets. In particular, we show how to reformulate the support functions of uncertainty sets represented in terms of matrix norms and cones. Our results provide extensions to known results from the literature. We also consider hard quadratic constraints; those that are convex in uncertain matrix-valued parameters. For the robust counterpart of such constraints we derive inner and outer tractable approximations. As application, we show how to construct a natural uncertainty set based on a statistical confidence set around a sample mean vector and covariance matrix and use this to provide a tractable reformulation of the robust counterpart of an uncertain portfolio optimization problem. We also apply the results of this paper to a norm approximation and a regression line problem.

## 1 Introduction

Many real-life optimization problems have parameters whose values are not exactly known. Let us consider an optimization problem containing the constraint

$$f(y, \zeta) \leq 0, \tag{1}$$

where  $y \in \mathbb{R}^n$  is the decision variable,  $\zeta \in \mathbb{R}^t$  is the parameter that is not known exactly, and  $f : \mathbb{R}^n \times \mathbb{R}^t \rightarrow \mathbb{R}$  is a continuous function. One way to deal with parameter uncertainty is Robust Optimization (RO), which enforces the constraints to hold for all uncertain parameter values in a user specified uncertainty set  $\mathcal{Z} \subseteq \mathbb{R}^t$ . More precisely, RO changes (1) into

$$f(y, \zeta) \leq 0, \quad \forall \zeta \in \mathcal{Z}. \tag{2}$$

This leads to a semi-infinite optimization problem, called the robust counterpart (RC), which is generally computationally intractable (see, e.g., Example 1.2.7 of the book [8]). A challenge in RO is to find a tractable, i.e., conic quadratic or semi-definite, reformulation of the RC. Tractability depends not only on the functions defining the constraint, i.e.,  $f(y, \zeta)$  in (1), but also on the uncertainty set  $\mathcal{Z}$ . For a linear constraint with linear uncertainty, where  $f(y, \zeta) = b(\zeta)^T y + c$ , with scalar  $c \in \mathbb{R}$  and affine  $b(\zeta) \in \mathbb{R}^n$ , there is a broad range of uncertainty sets for which the RC has a tractable reformulation, see [26].

---

<sup>\*</sup>Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, The Netherlands

<sup>†</sup>Corresponding author: a.marandi@tue.nl

<sup>‡</sup>CentER Extramural fellow, Tilburg University, The Netherlands

<sup>§</sup>Tilburg School of Economics and Management, Tilburg University, The Netherlands

An extension of the linear case that we consider in this paper is an uncertain quadratic constraint

$$y^T A(\Delta)y + b(\Delta)^T y + c \leq 0, \quad (3)$$

where  $A(\Delta) \in \mathbb{R}^{n \times n}$  and  $b(\Delta) \in \mathbb{R}^n$  are uncertain, and  $c \in \mathbb{R}$  is deterministic. We consider uncertain constraints in which the uncertainty in the parameters can be formulated in a matrix format, whereas the results in the literature are mainly for vector uncertainty. Throughout the paper, we use the notation  $\zeta$  in case of vector uncertainty and  $\Delta$  in case of matrix uncertainty. So, we consider the RC of (3):

$$y^T A(\Delta)y + b(\Delta)^T y + c \leq 0, \quad \forall \Delta \in \mathcal{Z}, \quad (4a)$$

where  $\Delta \in \mathbb{R}^{n \times n}$  (of the same dimension as  $A$ ) is the uncertain parameter belonging to the convex compact uncertainty set  $\mathcal{Z} \subseteq \mathbb{R}^{n \times n}$ , where  $A(\Delta) \in \mathbb{R}^{n \times n}$  and  $b(\Delta) \in \mathbb{R}^n$  are affine in  $\Delta$ ,  $A(\Delta)$  is positive semi-definite for all  $\Delta \in \mathcal{Z}$ , and where  $c \in \mathbb{R}$ .

An important optimization problem having constraints in the form (4a), is a *portfolio choice problem*, in which one tries to find an asset allocation that trades off a low risk against a high expected return. One can formulate a portfolio choice problem using the form (4a), where  $A(\Delta)$  is the covariance matrix and  $b(\Delta)$  is minus the vector of mean returns (possibly with a weight), respectively.

In addition to a quadratic constraint in the form (4a), we consider a conic quadratic constraint that is concave in the uncertain parameters in the form

$$\sqrt{y^T A(\Delta)y + b(\Delta)^T y + c} \leq 0, \quad \forall \Delta \in \mathcal{Z}, \quad (4b)$$

where  $A(\Delta)$ ,  $b(\Delta)$ ,  $c$ , and  $\mathcal{Z}$  are defined as above.

To the best of our knowledge, there are only a few papers treating the constraints in the forms (4). Moreover, the matrix  $A$  typically is given as an uncertain linear combination of some primitive matrices with vector uncertainty. For example, the authors in [24] study constraints in the form (3), where  $A$  is formulated as  $\sum_{i=1}^t \zeta_i A_i$  and  $\zeta = [\zeta_1, \dots, \zeta_t]^T \in \mathcal{Z} \subseteq \mathbb{R}^t$  is the uncertain parameter vector, for given positive semi-definite matrices  $A_i$ ,  $i = 1, \dots, t$ . They provide exact tractable reformulations of RCs for polyhedral and ellipsoidal uncertainty sets. The uncertainty set  $\mathcal{Z}$  that we consider in this paper is a matrix-valued one, which is not studied in [24]. The results in [29] are similar to the results in [24] when applied to a quadratic constraint in the form (3). In a more general setting, the authors in [25] introduce a dual problem to a general convex nonlinear robust optimization problem where the objective function and constraints are concave in the uncertain parameters, and provide conditions under which strong duality holds.

Except for the aforementioned papers, the focus in the literature remarkably is on the constraints in the forms

$$y^T A(\Delta)^T A(\Delta)y + b(\Delta)^T y + c \leq 0, \quad \forall \Delta \in \mathcal{Z}, \quad (5a)$$

$$\sqrt{y^T A(\Delta)^T A(\Delta)y + b(\Delta)^T y + c} \leq 0, \quad \forall \Delta \in \mathcal{Z}, \quad (5b)$$

where  $A(\Delta) \in \mathbb{R}^{m \times n}$  and  $b(\Delta) \in \mathbb{R}^n$  are affine in  $\Delta \in \mathcal{Z} \subseteq \mathbb{R}^{m \times n}$ , and  $\mathcal{Z}$  is a convex compact set. For example, the book [8] and papers [20] and [6] treat the constraints in the forms (5). The drawback of (5) is that the RC is, in general, (computationally) intractable, since the constraints are convex in the uncertain parameter  $\Delta$  (see, e.g., [34]).

It is worth mentioning that the key characteristic of the constraints in the forms (4) is that they are concave in  $\Delta$  with convex  $\mathcal{Z}$ . Constraints in the forms (5) can be formulated in terms of (4), for instance as follows:

$$y^T B(\bar{\Delta})y + b^T y + c \leq 0, \quad \forall \bar{\Delta} \in \bar{\mathcal{Z}},$$

$$\sqrt{y^T B(\bar{\Delta})y + b^T y + c} \leq 0, \quad \forall \bar{\Delta} \in \bar{\mathcal{Z}},$$

where  $B(\bar{\Delta}) = \bar{\Delta}$  and  $\bar{\mathcal{Z}} = \{\bar{\Delta} : \bar{\Delta} = A(\Delta)^T A(\Delta), \Delta \in \mathcal{Z}\}$ , but  $\bar{\mathcal{Z}}$  is not convex anymore, even not for a convex  $\mathcal{Z}$ .

On the one hand, the focus of the literature is on reformulating the RCs of constraints in the forms (5) with specific convex compact uncertainty sets, with applications specially in least-squares

problems. On the other hand, many applications that naturally contain constraints in the forms (4) with matrix-valued uncertainty sets have been left out from the literature. Some of the applications, in addition to portfolio choice problems, are the following ones.

- **Chance Constraint** [2, Chapter 1]: Consider a normally distributed random vector  $a \in \mathbb{R}^n$ . Let  $y \in \mathbb{R}^n$  be the vector of decision variables and  $c \in \mathbb{R}$  be a constant scalar, respectively. Then, the *chance constraint*  $\text{Prob}(a^T y + c \geq 0) \geq \alpha$  is equivalent to  $0 \geq z_\alpha \sqrt{y^T \Sigma y} - y^T \mu - c$ , where  $\alpha \in (0, 1)$ ,  $z_\alpha$  is the  $\alpha$  percentile of the standard normal distribution,  $\mu$  and  $\Sigma$  are the mean vector and covariance matrix of  $a$ , respectively. Usually,  $\mu$  and  $\Sigma$  are estimated based on historical data, which results in estimation inaccuracy. Since,  $\mu$  and  $\Sigma$  are uncertain, the inequality is of the format (4b).
- **Quadratic Approximations**: Many optimization methods, like (quasi) Newton and Sequential Quadratic Programming, use quadratic approximations of objective and constraint functions. For a twice differentiable function, this approximation can be taken using the second order truncated Taylor expansion, which requires calculating the gradient vector and the Hessian matrix. However, often the calculated gradients and Hessians are inaccurate, which make them uncertain. Therefore, if we apply methods, like the Newton method, to a convex optimization problem, then we could approximate it by a convex quadratic optimization problem, with an uncertain gradient vector and Hessian matrix.

The contribution of this paper is fourfold. First, we extend the results in [4], who consider vector uncertainty, to derive reformulations of the support functions of matrix-valued uncertainty sets. We derive explicit formulas for support functions of many choices of  $\mathcal{Z}$ , mostly of those given in terms of matrix norms and cones. We demonstrate that these derivations for support functions of matrix-valued uncertainty sets are also useful for a class of linear Adjustable Robust Optimization problems introduced in [9].

Second, we derive tractable formulations of the RCs of uncertain constraints in the forms (4), where  $A(\Delta)$  is positive semi-definite, with a general convex compact matrix-valued uncertainty set  $\mathcal{Z}$ , given in terms of its support function. In the literature only for very special uncertainty sets tractable formulations have been developed, whereas the results in this paper are for a broad range of uncertainty sets.

Third, we develop inner and outer tractable approximations of the RCs of constraints in the forms (5). We do this by substituting the quadratic term in the uncertain parameter with upper and lower bounds that are linear in the uncertain parameter and hence are in the forms (4). These results extend the literature in two ways. First of all, inner approximations for (5) have been proposed in the literature only for box or 2-norm type uncertainty sets [6, 10] while our approach is for a much broader range of uncertainty sets. Secondly, in this paper we also derive outer approximations. Hence, we obtain both a lower and an upper bound for the optimal value of the problem. In the literature mostly inner approximations are derived. We test these approximations on norm approximation problems as well as linear regression problems with budgeted-type uncertainty set, which could not be treated using the results in the literature. Our numerical experiments show that the obtained robust solutions outperform the nominal solutions.

Fourth, we show how to construct a natural uncertainty set consisting of the mean vector and the (vectorized) covariance matrix by using historical data and probabilistic confidence sets. This type of uncertainty sets is important for applications such as portfolio optimization problems. We prove for this type of sets that the support function is semi-definite representable, and provide a tractable reformulation of the robust counterpart of an uncertain portfolio optimization problem.

The remainder of the paper is organized as follows. Section 2 introduces notations and definitions that are used throughout the paper. In Section 3, we show how to derive computationally tractable expressions for the support functions of matrix-valued sets defined by matrix norms and cones, and several composition rules, including summations, intersections, Cartesian products of sets, convexification, linear transformations, and many more. In Section 4, we derive an exact tractable formulation for the RC of constraints in the forms (4) for a general convex compact uncertainty sets. In Section 5, we study constraints in the forms (5) with a general convex compact uncertainty set, and provide inner and outer approximations of the RCs. Section 6 is about constructing an uncertainty set using historical information and confidence sets. In Section 7, we apply the results of this paper to a portfolio choice, a norm approximation, and a regression line problem.

This paper contains four appendices. Appendix A contains the proofs of lemmas and propositions not presented in the main text. The second appendix contains simple illustrative examples for the results in Section 4. In Appendix C, we show how one can check assumptions needed in Section 5 to derive the approximations. Finally, Appendix D contains a heuristic method to find worst-case scenarios, which are used in the numerical experiments to check the quality of the solutions obtained using the inner and outer approximations proposed in Section 5.

## 2 Preliminaries

In this section, we introduce the notations and definitions we use throughout the paper. We denote by  $S_n$  the set of all  $n \times n$  symmetric matrices, and by  $S_n^+$  its subset of all positive semi-definite matrices. For  $A, B \in \mathbb{R}^{n \times n}$ , the notations  $A \succeq B$  and  $A \succ B$  are used when  $A - B \in S_n^+$  and  $A - B \in \text{int}(S_n^+)$ , respectively, where  $\text{int}(S_n^+)$  denotes the interior of  $S_n^+$ . We denote by  $\text{trace}(A)$  the trace of  $A$ . For  $A, B \in \mathbb{R}^{n \times m}$ , we set  $\text{vec}(A) := [A_{11}, \dots, A_{1m}, \dots, A_{n1}, \dots, A_{nm}]^T$ , and hence,  $\text{trace}(AB^T) = \text{vec}(A)^T \text{vec}(B)$ . For symmetric matrices  $A, B \in S_n$ , we set  $\text{svec}(A) := [A_{11}, \sqrt{2}A_{12}, \dots, \sqrt{2}A_{1n}, A_{22}, \dots, \sqrt{2}A_{(n-1)n}, A_{nn}]^T$ , and hence,  $\text{trace}(AB) = \text{svec}(A)^T \text{svec}(B)$ . Additionally, to represent a vector  $d \in \mathbb{R}^n$  by its components, we use  $[d_i]_{i=1, \dots, n}$ . Also, we denote the zero matrix in  $\mathbb{R}^{n \times m}$  and identity matrix in  $S_n$  by  $0_{n \times m}$  and  $I_n$ , respectively. Moreover, for matrices  $A, B \in \mathbb{R}^{n \times m}$ , we denote the Hadamard product by  $A \circ B$ ; i.e., for any  $i = 1, \dots, n$  and  $j = 1, \dots, m$ , we have  $(A \circ B)_{ij} = A_{ij}B_{ij}$ .

We denote the singular values of a matrix  $A \in \mathbb{R}^{m \times n}$  with rank  $r$  by  $\sigma_1(A) \geq \dots \geq \sigma_r(A) > 0$ . For a vector  $x \in \mathbb{R}^n$ , the Euclidean norm is denoted by  $\|x\|_2$ . We use the following matrix norms in this paper:

**Frobenius norm:**  $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$ ;

$l_1$  **norm:**  $\|A\|_1 = \sum_{i=1}^m \sum_{j=1}^n |A_{ij}|$ ;

$l_\infty$  **norm:**  $\|A\|_\infty = \max_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} |A_{ij}|$ ;

**spectral norm:**  $\|A\|_{2,2} = \sup_{\|x\|_2=1} \|Ax\|_2$ ;

**trace (nuclear) norm:**  $\|A\|_\Sigma = \sigma_1(A) + \dots + \sigma_r(A)$ ;

**dual norm:** For a general matrix norm  $\|\cdot\|$ , its dual norm is defined as  $\|A\|^* = \max_{\|B\|=1} \text{trace}(B^T A)$ .

**Remark 1.** Let  $\|\cdot\|$  be a general vector norm. Then a matrix norm can be defined as  $\|\text{vec}(A)\|$  for a matrix  $A \in \mathbb{R}^{m \times n}$ . Frobenius,  $l_1$ , and  $l_\infty$  norms are examples of this type of matrix norms.  $\square$

The following lemma provides the exact formulations of the dual norms corresponding to the matrix norms defined above.

**Lemma 1.** [28, Section 5.6]

(a)  $\|A\|_F^* = \|A\|_F = \|\text{vec}(A)\|_2$ ; (b)  $\|A\|_1^* = \|A\|_\infty$ ; (c)  $\|A\|_\Sigma^* = \|A\|_{2,2} = \sigma_1(A)$ .  $\square$

In the rest of this section, we recall some definitions related to optimization.

**Definition 1.** Let  $\mathcal{Y}$  be a set determined by constraints in a variable  $y$ . A set  $\mathcal{S}$  determined by constraints in the variable  $y$  and additional variable  $x$ , is an inner approximation of  $\mathcal{Y}$ , if  $(x, y) \in \mathcal{S} \Rightarrow y \in \mathcal{Y}$ . A set  $\mathcal{S}$  is an outer approximation if  $y \in \mathcal{Y} \Rightarrow \exists x : (x, y) \in \mathcal{S}$ .  $\square$

In [8] the inner approximation is called safe approximation.

**Definition 2.** For a convex set  $\mathcal{Z}$ , the support function  $\delta_{\mathcal{Z}}^*(\cdot)$  is defined as follows:

$$\begin{aligned} \text{if } \mathcal{Z} \subseteq \mathbb{R}^n, \quad & \delta_{\mathcal{Z}}^*(u) := \sup_{b \in \mathcal{Z}} \{u^T b\}, \\ \text{if } \mathcal{Z} \subseteq \mathbb{R}^{m \times n}, \quad & \delta_{\mathcal{Z}}^*(W) := \sup_{A \in \mathcal{Z}} \{\text{trace}(AW^T)\}, \\ \text{if } \mathcal{Z} \subseteq \mathbb{R}^{m \times n} \times \mathbb{R}^n, \quad & \delta_{\mathcal{Z}}^*(W, u) := \sup_{(A, b) \in \mathcal{Z}} \{\text{trace}(AW^T) + u^T b\}, \end{aligned}$$

where  $W \in \mathbb{R}^{m \times n}$ ,  $u \in \mathbb{R}^n$ .  $\square$

**Definition 3.** Let

$$\mathcal{Z} = \left\{ \Delta \in \mathbb{R}^{m \times n} : \begin{array}{ll} \text{trace}(C^i \Delta) + q^i = 0, & i = 1, \dots, I, \\ h_\ell(\Delta) \leq 0, & \ell = 1, \dots, L, \\ g_k(\Delta) \preceq 0_{p \times p}, & k = 1, \dots, K \end{array} \right\}, \quad (6)$$

where  $p$  is a positive integer,  $C^i \in \mathbb{R}^{m \times n}$ ,  $i = 1, \dots, I$ ,  $q \in \mathbb{R}^I$ , and  $h_\ell : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ ,  $\ell = 1, \dots, L$ , and  $g_k : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{p \times p}$ ,  $k = 1, \dots, K$ , are convex continuous functions. Slater condition is satisfied for  $\mathcal{Z}$  if there exists  $\bar{\Delta} \in \mathcal{Z}$  such that  $h_\ell(\bar{\Delta}) < 0$ , for any  $\ell = 1, \dots, L$ , and  $g_k(\bar{\Delta}) \prec 0_{t \times t}$ , for any  $k = 1, \dots, K$ . We call  $\bar{\Delta}$  a Slater point.  $\square$

### 3 Support functions for matrix-valued uncertainty sets

The importance of the support functions of the vector-valued sets in the area of Robust Optimization has been highlighted by [4], who show how to derive explicit formulas of the support functions. Support functions of the matrix-valued sets, however, are not studied in the literature despite their applicabilities in defining uncertainty in linear optimization [13], quadratic optimization [20], and semi-definite optimization problems [19]. For instance, let us consider an Adjustable Robust Linear Optimization (ARO) problem, introduced in [9]:

$$\begin{aligned} \min_{x \in \mathbb{R}^t} c^T x + \max_{\Delta \in \mathcal{Z}} \min_{y(\Delta) \in \mathbb{R}^n} d^T y(\Delta) \\ \text{s.t. } \Delta x + B y(\Delta) \leq h, \\ y(\Delta) \geq 0, \end{aligned} \quad (7)$$

where  $x \in \mathbb{R}^t$  is a “here-and-now” decision,  $\Delta \in \mathbb{R}^{m \times t}$  is the uncertain parameter,  $\mathcal{Z} \subseteq \mathbb{R}^{m \times t}$  is a convex compact set,  $y(\cdot) \in \mathbb{R}^n$  is a “wait-and-see” variable,  $c \in \mathbb{R}^t$ ,  $B \in \mathbb{R}^{m \times n}$ ,  $d \in \mathbb{R}^n$ , and  $h \in \mathbb{R}^m$ . A typical approach to approximate the ARO problem (7) is to restrict the “wait-and-see” variable  $y(\Delta)$  to be affine in the uncertainty parameter. In other words, (7) is approximated by

$$\begin{aligned} \min_{\substack{\tau \in \mathbb{R}, x \in \mathbb{R}^t \\ V^i \in \mathbb{R}^{m \times n} \\ u^i \in \mathbb{R}}} c^T x + \tau \\ \text{s.t. } \tau \geq \sum_{i=1}^n d_i (\text{trace}(V^i \Delta) + u^i), \quad \forall \Delta \in \mathcal{Z}, \\ \Delta x + \left[ \sum_{i=1}^n B_{ji} (\text{trace}(V^i \Delta) + u^i) \right]_{j=1, \dots, m} \leq h, \quad \forall \Delta \in \mathcal{Z}, \\ [\text{trace}(V^i \Delta) + u^i]_{i=1, \dots, n} \geq 0, \quad \forall \Delta \in \mathcal{Z}, \end{aligned} \quad (8)$$

which is the robust counterpart of an uncertain linear optimization problem where the uncertain parameters appear in all constraints. Problem (8) can be solved efficiently if the support function of  $\mathcal{Z}$  has a tractable reformulation. In this section, we focus on deriving explicit tractable formulations of the support functions of matrix-valued sets.

In the following lemma we provide equivalent formulations of the support functions of the sets constructed using standard composition rules.

**Lemma 2.** Let  $U \in \mathbb{R}^{n \times n}$ .

- (i) Let  $\mathcal{Z} = \left\{ \Delta \in \mathbb{R}^{n \times n} : \text{vec}(\Delta) \in U \subseteq \mathbb{R}^{n^2} \right\}$ . Then  $\delta_{\mathcal{Z}}^*(U) = \delta_U^*(\text{vec}(U))$ .
- (ii) Let  $\Delta^1, \dots, \Delta^k \in \mathbb{R}^{n \times n}$  be given. Also, let  $\mathcal{Z} = \left\{ \sum_{i=1}^k \zeta_i \Delta^i : \zeta \in U \subseteq \mathbb{R}^k \right\}$ . Then,  $\delta_{\mathcal{Z}}^*(U) = \delta_U^* \left( [\text{trace}(\Delta^i U^T)]_{i=1, \dots, k} \right)$ .
- (iii) Let  $L \in \mathbb{R}^{n \times t}$  and  $R \in \mathbb{R}^{s \times n}$  be given, and  $\mathcal{Z} = \{ L \Delta R : \Delta \in U \subseteq \mathbb{R}^{t \times s} \}$ . Then  $\delta_{\mathcal{Z}}^*(U) = \delta_U^*(L^T U R^T)$ .

(iv) Let  $L \in \mathbb{R}^{n \times n}$  be given, and  $\mathcal{Z} = \{\Delta : L \circ \Delta \in \mathcal{U} \subseteq \mathbb{R}^{n \times n}\}$ . Then,

$$\delta_{\mathcal{Z}}^*(U) = \begin{cases} \delta_{\mathcal{U}}^*(U \circ L^\dagger) & \text{if } U_{ij} = 0 \text{ for any } i, j = 1, \dots, n, \text{ such that } L_{ij} = 0, \\ +\infty & \text{if } U_{ij} \neq 0 \text{ for some } i, j = 1, \dots, n, \text{ such that } L_{ij} = 0, \end{cases}$$

where for any  $i, j = 1, \dots, n$ ,

$$L_{ij}^\dagger = \begin{cases} \frac{1}{L_{ij}} & \text{if } L_{ij} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(v) Let  $\mathcal{Z}_i \subseteq \mathbb{R}^{n \times n}$ ,  $i = 1, \dots, k$ , and let  $\mathcal{Z} = \sum_{i=1}^k \mathcal{Z}_i$  be the Minkowski sum. Then  $\delta_{\mathcal{Z}}^*(U) = \sum_{i=1}^k \delta_{\mathcal{Z}_i}^*(U)$ .

(vi) Let  $\mathcal{Z}_i \subseteq \mathbb{R}^{n \times n}$ ,  $i = 1, \dots, k$ , be in the form (6) and have a common Slater point. Also, let  $\mathcal{Z} = \bigcap_{i=1}^k \mathcal{Z}_i$ . Then  $\delta_{\mathcal{Z}}^*(U) = \min_{U^i \in \mathbb{R}^{n \times n}} \left\{ \sum_{i=1}^k \delta_{\mathcal{Z}_i}^*(U^i) : \sum_{i=1}^k U^i = U \right\}$ .

(vii) Let  $\mathcal{Z}_i \subseteq \mathbb{R}^{n_i \times n_i}$ ,  $U_i \in \mathbb{R}^{n_i \times n_i}$ ,  $i = 1, \dots, k$ , and  $\mathcal{Z} = \{\Delta = (\Delta_1, \dots, \Delta_k) : \Delta_i \in \mathcal{Z}_i, i = 1, \dots, k\}$ . Then we have  $\delta_{\mathcal{Z}}^*((U_1, \dots, U_k)) = \sum_{i=1}^k \delta_{\mathcal{Z}_i}^*(U_i)$ .

(viii) Let  $\mathcal{Z}_i \subseteq \mathbb{R}^{n \times n}$ ,  $i = 1, \dots, k$ , be convex and  $\mathcal{Z} = \text{conv}(\bigcup_{i=1}^k \mathcal{Z}_i)$  be the convex hull. Then  $\delta_{\mathcal{Z}}^*(U) = \max_{i=1, \dots, k} \delta_{\mathcal{Z}_i}^*(U)$ .

*Proof.* Proof. Appendix A.1. □

Lemma 2 shows how we can derive the support functions of sets constructed using different composition rules without being restricted to vector-valued sets in contrast with the results in [4], which hold for vector-valued sets.

In the next lemma we derive explicit tractable reformulations of the support functions of matrix-valued uncertainty sets defined by matrix norms and the cone of positive semi-definite matrices.

**Lemma 3.** Let  $U \in \mathbb{R}^{n \times n}$ .

- (a) Let  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{n \times n} : \|\Delta\| \leq \rho\}$ , where  $\|\cdot\|$  is a general matrix norm. Then  $\delta_{\mathcal{Z}}^*(U) = \rho \|U\|^*$ .
- (b) Let  $\mathcal{Z} = \{\Delta : \Delta^l \preceq \Delta \preceq \Delta^u\}$ , where  $\Delta^l, \Delta^u \in S_n$  are given such that  $\Delta^u - \Delta^l \succ 0_{n \times n}$ . Then

$$\delta_{\mathcal{Z}}^*(U) = \min_{\Lambda_1, \Lambda_2} \left\{ \text{trace}(\Delta^u \Lambda_2) - \text{trace}(\Delta^l \Lambda_1) : \Lambda_2 - \Lambda_1 = \frac{U + U^T}{2}, \Lambda_1, \Lambda_2 \succeq 0_{n \times n} \right\}.$$

*Proof.* Proof. (a) This follows directly from the definition of the dual norm.

(b) This follows directly from conic duality (see Appendix A.2). □

Special cases of the uncertainty sets studied in Lemma 3 have been considered in the literature. The uncertainty set constructed using the *Frobenius norm* is considered in [20] for the constraints in the form (5b). Also, the authors of [36] construct an uncertainty set for the covariance matrix using the *Frobenius norm*. The constraints in the forms (5) with uncertainty set defined by the *spectral norm* is treated in Chapter 6 of [8]. Furthermore, the uncertainty set that we considered in Lemma 3(b) is constructed in [35] for covariance matrices. Besides, the authors of [18] construct an uncertainty set for the mean vector and covariance matrix, which can be formulated as an intersection of two sets that are considered in Lemma 3(b).

It is known that the  $l_1$  and  $l_\infty$  norms are linear representable and the *Frobenius norm* is conic quadratic representable. The following lemma shows that the *spectral* and *trace norms* are semi-definite representable.

**Lemma 4.** Let  $U \in \mathbb{R}^{n \times n}$  and  $\rho \geq 0$ .

- (i)  $\|U\|_\Sigma \leq \rho$  if and only if there exist matrices  $Y \in \mathbb{R}^{n \times n}$  and  $Z \in \mathbb{R}^{n \times n}$  such that

$$\begin{bmatrix} Y & U \\ U^T & Z \end{bmatrix} \succeq 0_{2n \times 2n}, \quad \text{trace}(Y) + \text{trace}(Z) \leq 2\rho.$$

(ii)  $\|U\|_{2,2} \leq \rho$  if and only if  $\begin{bmatrix} \rho^2 I_n & U \\ U^T & I_n \end{bmatrix} \succeq 0_{2n \times 2n}$ .

*Proof.* Proof. (i) See, e.g., Lemma 1 in [22].

(ii) See, e.g., Example 8 in [4], or Appendix A.3.  $\square$

Hitherto, we have shown how to derive tractable reformulations of the support functions of matrix-valued uncertainty sets. In the next section, we show how such reformulations can be used in tractably reformulating the RC of an uncertain quadratic constraint in the forms (4).

## 4 Tractable reformulation of Robust Quadratic Optimization problems that are concave in the uncertain parameters

In this section, we assume that  $A(\Delta) = A + \Delta$ . We emphasize that this assumption can be made without loss of generality for  $A(\Delta)$  that is affine on  $\Delta$ , because of Lemma 2.(iii). Moreover, from here on in the paper, we assume that the uncertainty set  $\mathcal{Z}$  is defined as in (6) and satisfies the Slater condition. The next theorem, which is the main theorem in this section, provides reformulations of the RCs of constraints in the forms (4), and asserts that their tractabilities only depend on the uncertainty sets.

**Theorem 1.** *Let  $\mathcal{Z} \subset \mathbb{R}^{n \times n}$  be a convex, compact set. Also, let  $\bar{A} \in \mathbb{R}^{n \times n}$ ,  $a, \bar{b} \in \mathbb{R}^n$ , and  $c \in \mathbb{R}$  be given. For any  $\Delta \in \mathcal{Z} \subseteq \mathbb{R}^{n \times n}$ , let  $A(\Delta) = \bar{A} + \Delta$ ,  $b(\Delta) = \bar{b} + \Delta a$  ( $m = n$  in (4)). Assume that  $A(\Delta)$  is positive semi-definite (PSD), for all  $\Delta \in \mathcal{Z}$ , and that for a Slater point  $\bar{\Delta}$ ,  $A(\bar{\Delta})$  is positive definite. Then:*

(I)  $y \in \mathbb{R}^n$  satisfies (4a) if and only if there exists  $W \in \mathbb{R}^{n \times n}$  satisfying the convex system

$$\text{trace}(\bar{A}W) + \bar{b}^T y + c + \delta_{\mathcal{Z}}^*(W + ya^T) \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad (9)$$

(II)  $y \in \mathbb{R}^n$  satisfies (4b) if and only if there exist  $W \in \mathbb{R}^{n \times n}$  and  $\eta \in \mathbb{R}$  satisfying the convex system

$$\text{trace}(\bar{A}W) + \bar{b}^T y + c + \delta_{\mathcal{Z}}^*(W + ya^T) + \frac{\eta}{4} \leq 0, \quad \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad (10)$$

*Proof.* Proof. To prove this theorem we use the same line of reasoning as in Theorem 2 in [4]. For any  $\Delta \in \mathcal{Z}$ , it is clear that  $\frac{A(\Delta) + A(\Delta)^T}{2} \succeq 0_{n \times n}$  due to positive semi-definiteness of  $A(\Delta)$ . Also,  $y^T A(\Delta)y = y^T \frac{A(\Delta) + A(\Delta)^T}{2} y$  for any  $y \in \mathbb{R}^n$ , and  $\Delta \in \mathcal{Z}$ . We replace  $y^T A(\Delta)y$  by  $y^T \frac{A(\Delta) + A(\Delta)^T}{2} y$  in constraints (4).

(I) Let  $\mathcal{U} = \left\{ \left( \frac{A(\Delta) + A(\Delta)^T}{2}, b(\Delta) \right) : \Delta \in \mathcal{Z} \right\}$ . It is clear that  $y \in \mathbb{R}^n$  satisfies (4a) if and only if  $F(y) := \max_{(B,b) \in \mathcal{U}} \{y^T B y + b^T y + c\} \leq 0$ . Setting

$$\delta_{\mathcal{U}}(B, b) = \begin{cases} 0 & \text{if } (B, b) \in \mathcal{U}, \\ +\infty & \text{otherwise,} \end{cases}$$

we have  $F(y) = \max_{\substack{B \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \{y^T B y + b^T y + c - \delta_{\mathcal{U}}(B, b)\}$ . Since  $B \succeq 0_{n \times n}$  for all  $B \in \mathcal{U}$ , and  $\frac{A(\bar{\Delta}) + A(\bar{\Delta})^T}{2}$  is positive definite and lies in the relative interior of  $\mathcal{U}$ , specialization of Theorem 4.4.3 in [15] to  $\mathbb{R}^{n \times n}$  implies that  $F(y) \leq 0$  is equivalent to the existence of  $W \in \mathbb{R}^{n \times n}$  and  $u \in \mathbb{R}^n$ , such that

$$\delta_{\mathcal{U}}^*(W, u) - \inf_{\substack{A \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \{ \text{trace}(AW^T) + u^T b - (y^T A y + b^T y + c) \} \leq 0, \quad (11)$$

where  $\delta_{\mathcal{U}}^*(\cdot)$  is the support function of the set  $\mathcal{U}$ . It follows from Definition 2 that

$$\begin{aligned}
\delta_{\mathcal{U}}^*(W, u) &= \sup_{(B, b) \in \mathcal{U}} \{ \text{trace}(BW^T) + u^T b \} = \sup_{\Delta \in \mathcal{Z}} \left\{ \text{trace} \left( \frac{A(\Delta) + A(\Delta)^T}{2} W^T \right) + u^T b(\Delta) \right\} \\
&= \sup_{\Delta \in \mathcal{Z}} \left\{ \text{trace} \left( (\bar{A} + \Delta) \left( \frac{W + W^T}{2} \right) \right) + u^T (\bar{b} + \Delta a) \right\} \\
&= \text{trace} \left( \bar{A} \left( \frac{W + W^T}{2} \right) \right) + u^T \bar{b} + \sup_{\Delta \in \mathcal{Z}} \left\{ \text{trace} \left( \Delta \left( \frac{W + W^T}{2} \right) \right) + u^T \Delta a \right\} \\
&= \text{trace} \left( \bar{A} \left( \frac{W + W^T}{2} \right) \right) + u^T \bar{b} + \sup_{\Delta \in \mathcal{Z}} \left\{ \text{trace} \left( \Delta \left( \left( \frac{W + W^T}{2} \right) + a u^T \right) \right) \right\} \\
&= \text{trace} \left( \bar{A} \left( \frac{W + W^T}{2} \right) \right) + u^T \bar{b} + \delta_{\mathcal{Z}}^* \left( \left( \frac{W + W^T}{2} \right) + a u^T \right). \tag{12}
\end{aligned}$$

Also, we have

$$\begin{aligned}
&\inf_{\substack{A \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \{ \text{trace}(AW^T) + u^T b - (y^T A y + b^T y + c) \} \\
&= \inf_{\substack{A \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \{ \text{trace}(AW) + u^T b - (y^T A y + b^T y + c) \} \\
&= -c + \inf_{\substack{A \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \{ \text{trace}(A(W - yy^T)) + b^T(u - y) \} = \begin{cases} -c & W - yy^T \succeq 0_{n \times n}, u = y \\ -\infty & \text{otherwise.} \end{cases} \tag{13}
\end{aligned}$$

So, the fact that  $W \succeq 0_{n \times n}$  implies  $\frac{W+W^T}{2} = W$ , and the Schur Complement Lemma (see, e.g., Appendix A.5.5 in [16]), (12), and (13) result in (9).

(II) Similar to the proof of part (I) we have  $y \in \mathbb{R}^n$  satisfies (4b) if and only if there exists  $W \in \mathbb{R}^{n \times n}$  such that

$$\delta_{\mathcal{U}}^*(W, u) - \inf_{\substack{A \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \left\{ \text{trace}(AW^T) + u^T b - \left( \sqrt{y^T A y} + b^T y + c \right) \right\} \leq 0. \tag{14}$$

Analogous to the result in Section 3.4 in [25],

$$\inf_{\substack{A \succeq 0_{n \times n} \\ b \in \mathbb{R}^n}} \left\{ \text{trace}(AW^T) + u^T b - \left( \sqrt{y^T A y} + b^T y + c \right) \right\} = -c - \inf_{\eta} \left\{ \frac{\eta}{4} : u = y, \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1} \right\}.$$

So, (14) is equivalent to

$$\delta_{\mathcal{U}}^*(W, u) + c + \inf_{\eta \in \mathbb{R}} \left\{ \frac{\eta}{4} : \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1} \right\} \leq 0. \tag{15}$$

In (15), the infimum is taken over a closed lower bounded set, since  $\eta \geq 0$ . Hence,  $W \in \mathbb{R}^{n \times n}$  and  $y \in \mathbb{R}^n$  satisfies (15) if and only if there exists  $\eta \in \mathbb{R}$  such that

$$\text{trace}(W\bar{A}^T) + \bar{b}^T y + \delta_{\mathcal{Z}}^*(W + ya^T) + c + \frac{\eta}{4} \leq 0, \quad \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1},$$

which completes the proof.  $\square$

One of the assumptions in Theorem (1) is that  $A(\Delta)$  is positive semi-definite for all  $\Delta \in \mathcal{Z}$ . This assumption is needed to guarantee convexity of the constraint. Even though checking this assumption for a general uncertainty set is intractable (Section 8.2 in [8]), there are cases for which this assumption holds. An example is when  $A(\Delta)$  is a covariance matrix, which is estimated, e.g., based on historical data. Another example is when  $A(\Delta)$  is the Laplacian matrix of a weighted graph, where the weights are uncertain. In these examples,  $A(\Delta)$  by construction is positive semi-definite for all possible values of the uncertain parameter  $\Delta$ . Besides the aforementioned examples,



it is clear that if  $\bar{A}$  is positive semi-definite and  $\mathcal{Z} \subseteq S_n^+$ , then  $A(\Delta)$  is positive semi-definite for all  $\Delta \in \mathcal{Z}$ .

Next to the cases mentioned above, Theorem 8.2.3 in [8] provides a tractable method to check this assumption for a specific class of uncertainty sets. In the following lemma, we mention a simplified version of this theorem.

**Lemma 5.** (Theorem 8.2.3 in [8]) *Let  $\mathcal{Z} = \{\Delta : \|\Delta\|_{2,2} \leq \rho\} \subset \mathbb{R}^{n \times n}$ . Then, for a given  $\bar{A} \in \mathbb{R}^{n \times n}$ , we have that  $\bar{A} + \Delta \succeq 0_{n \times n}$  for any  $\Delta \in \mathcal{Z}$  if and only if  $\bar{A} - \rho I_n \succeq 0_{n \times n}$ .  $\square$*

## Illustrative examples

In the rest of this section, we derive tractable reformulations of RCs for some natural uncertain convex quadratic and conic quadratic constraints. For brevity of exposition, we provide in Appendix B the tractable reformulation of an uncertain convex quadratic constraint where the uncertainty set is defined by the Frobenius norm, as well as an uncertain conic quadratic constraint where the uncertainty set is similar to the one proposed in [18].

The following example is for constraints in the form (3) with vector uncertainty.

**Example 1.** *Consider*

$$y^T A(\zeta)y + b(\zeta)^T y + c \leq 0, \quad \forall \zeta \in \mathcal{Z}, \quad (16)$$

where  $A(\zeta) = \bar{A} + \sum_{i=1}^t \zeta_i A^i$ ,  $b(\zeta) = \bar{b} + \sum_{i=1}^t \zeta_i b^i$ ,  $(A^i, b^i) \in \mathbb{R}^{n \times n} \times \mathbb{R}^n$  is given,  $i = 1, \dots, t$ . This constraint is considered in [24], where  $\bar{A}$  and  $A^i$ ,  $i = 1, \dots, t$ , are positive semi-definite and where  $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2$ , for some polyhedral or ellipsoidal sets  $\mathcal{Z}_1 \subseteq \mathbb{R}^m$  and  $\mathcal{Z}_2 \subseteq \mathbb{R}^{t-m}$ ,  $m \in \mathbb{N}$ , with  $A^i = 0_{n \times n}$ ,  $i = 1, \dots, m$ , and  $b^i = 0_{n \times 1}$ ,  $i = m+1, \dots, t$  (uncertainty in  $A$  is independent of the uncertainty in  $b$ ). In this example we show how using the results of Section 3 can extend the results of [24] for general uncertainty sets, where  $A(\zeta)$  is positive semi-definite for all  $\zeta \in \mathcal{Z}$ . Let

$$\mathcal{Z} = \left\{ \zeta \in \mathbb{R}^t : \begin{array}{l} C\zeta + q = 0_{I \times 1}, \\ h_\ell(\zeta) \leq 0, \quad \ell = 1, \dots, L, \\ g_k(\zeta) \leq 0_{p \times p}, \quad k = 1, \dots, K \end{array} \right\},$$

where  $p$  is a positive integer,  $C \in \mathbb{R}^{I \times t}$ ,  $q \in \mathbb{R}^I$ , and where  $h_\ell : \mathbb{R}^t \rightarrow \mathbb{R}$ ,  $\ell = 1, \dots, L$ , and  $g_k : \mathbb{R}^t \rightarrow \mathbb{R}^{p \times p}$ ,  $k = 1, \dots, K$ , are convex continuous functions. First, we assume that  $\mathcal{Z} \subseteq \mathbb{R}_+^t$ , where  $\mathbb{R}_+^t$  denotes the nonnegative orthant of  $\mathbb{R}^t$ . Also, we assume that  $\bar{A}$  and  $A_i$ ,  $i = 1, \dots, t$ , are positive semi-definite, and there is a Slater point in  $\mathcal{Z}$  for which  $A(\zeta)$  is positive definite. In this case,  $y \in \mathbb{R}^n$  satisfies (16) if and only if there exists  $v \in \mathbb{R}^t$  such that

$$y^T \bar{A}y + \delta_{\mathcal{Z}}^*(v) + \bar{b}^T y + c \leq 0, \quad v \geq \left[ y^T A^i y + b^{i^T} y \right]_{i=1, \dots, t}, \quad (17)$$

whose proof can be found in Appendix A.4. It is clear that in this case  $A(\zeta)$  is positive semi-definite for all  $\zeta \in \mathcal{Z}$ . Moreover, as mentioned in Remark 2 of [24], if  $b^i = 0$ ,  $i = 1, \dots, t$ , then, for a general uncertainty set  $\mathcal{Z}$ ,  $y \in \mathbb{R}^n$  satisfies (16) if and only if

$$y^T \left( \bar{A} + \sum_{i=1}^t \zeta_i A_i \right) y + \bar{b}^T y + c \leq 0, \quad \forall \zeta \in \bar{\mathcal{Z}},$$

where  $\bar{\mathcal{Z}} = \mathcal{Z} \cap \{\zeta : \zeta \geq 0_{t \times 1}\}$  and  $\bar{\mathcal{Z}} \neq \emptyset$ . If the uncertainty set  $\mathcal{Z}$  is a polyhedron, then a tractable RC is provided in [24]. For other types of uncertainty sets, like ellipsoidal uncertainty sets, deriving tractable RCs is achievable using the results in Sections 3 and 4. Let  $\bar{\mathcal{Z}} \neq \emptyset$ . Then  $y \in \mathbb{R}^n$  satisfies (16) if and only if

$$y^T \bar{A}y + \delta_{\bar{\mathcal{Z}}}^*(v) + \bar{b}^T y + c \leq 0, \quad v \geq \left[ y^T A^i y \right]_{i=1, \dots, t}.$$

This is an extension of the results of [24], since there is a broad range of uncertainty sets for which the support functions have tractable reformulations.

Now, for a general case where the uncertainty in  $A$  and  $b$  can be dependent, if  $A(\zeta)$  is positive semi-definite for all  $\zeta \in \mathcal{Z}$ , and positive definite for a Slater point, then by Theorem 1(I),  $y$  satisfies (16) if and only if there exists  $W \in \mathbb{R}^{n \times n}$  such that

$$\text{trace}(\bar{A}W) + \bar{b}^T y + \delta_{\mathcal{Z}}^* \left( \left[ \text{trace}(A^i W) + b^{i^T} y \right]_{i=1, \dots, t} \right) + c \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad \square$$

In Section 6, we derive a natural uncertainty set for a vector that consists of the mean vector and the vectorized covariance matrix. This type of uncertainty set can be used in different applications, such as portfolio choice problems. In the following example we derive a tractable reformulation of a quadratic constraint with an uncertainty set similar to the one constructed in Section 6.

**Example 2.** Consider the uncertain quadratic constraint

$$y^T \Delta y + \zeta^T y + c \leq 0, \quad \forall \begin{pmatrix} \zeta \\ \text{svec}(\Delta) \end{pmatrix} \in \mathcal{Z}$$

where  $\mathcal{Z} = \mathcal{Z}_1 \cap \mathcal{Z}_2$ , and

$$\mathcal{Z}_1 = \left\{ \begin{pmatrix} \zeta \\ \text{svec}(\Delta) \end{pmatrix} = B\nu : \|\nu\|_2 \leq \rho, \nu \in \mathbb{R}^{\frac{n^2+3n}{2}} \right\}, \quad \mathcal{Z}_2 = \left\{ \begin{pmatrix} \zeta \\ \text{svec}(\Delta) \end{pmatrix} : \zeta \in \mathbb{R}^n, \Delta \in S_n^+ \right\},$$

for some invertible  $B \in \mathbb{R}^{\frac{n^2+3n}{2} \times \frac{n^2+3n}{2}}$ ,  $\rho > 0$ . For a fixed  $W \in S_n$ , by Lemma 2(vi), and Example 4 in [4],

$$\delta_{\mathcal{Z}}^* \begin{pmatrix} u \\ \text{svec}(W) \end{pmatrix} = \begin{cases} \min_{\substack{u^1, u^2 \\ W^1, W^2}} \rho \left\| B^T \begin{pmatrix} u^1 \\ \text{svec}(W^1) \end{pmatrix} \right\|_2 + \delta_{\mathcal{Z}_2}^* \begin{pmatrix} u^2 \\ \text{svec}(W^2) \end{pmatrix} \\ \text{s.t. } u^1 + u^2 = u, \quad W^1 + W^2 = W, \quad W^1, W^2 \in S_n. \end{cases}$$

Similar to the proofs of Lemmas 2(i) and 2(vii), we have

$$\delta_{\mathcal{Z}}^* \begin{pmatrix} u \\ \text{svec}(W) \end{pmatrix} = \min_{W^1} \left\{ \rho \left\| B^T \begin{pmatrix} u \\ \text{svec}(W^1) \end{pmatrix} \right\|_2 : W^1 \succeq W \right\}. \quad (18)$$

It is easy to show that there exists a Slater point in  $\mathcal{Z}$ . Hence,  $y \in \mathbb{R}^n$  satisfies (2) if and only if there exists  $W \in S_n^+$  that satisfies  $\rho \left\| B^T \begin{pmatrix} u \\ \text{svec}(W) \end{pmatrix} \right\|_2 + c \leq 0$ ,  $\begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}$ .  $\square$

## 5 Tractable inner and outer approximations of Robust Quadratic Optimization problems that are convex in the uncertain parameters

In this section, we provide inner and outer approximations of the RCs of constraints in the forms (5) by replacing the quadratic term in the uncertain parameter with suitable upper and lower bounds. We assume that  $A(\Delta) = \bar{A} + \Delta$ ,  $b(\Delta) = \bar{b} + D\Delta a$ , for given  $D \in \mathbb{R}^{n \times m}$ , full column-rank matrix  $\bar{A} \in \mathbb{R}^{m \times n}$ , vectors  $\bar{b} \in \mathbb{R}^n$ ,  $a \in \mathbb{R}^n$ , and scalar  $c \in \mathbb{R}$ , and that  $\Delta \in \mathcal{Z} \subseteq \mathbb{R}^{m \times n}$ , where  $\mathcal{Z}$  is a convex and compact set in the form (6) containing  $0_{m \times n}$  as a Slater point. Here, we list all assumptions on the constraints in the forms (5a) and (5b) that we will make in this section, and use some of them in each theorem.

**Assumption:**

- (A) there exists  $\Omega > 0$  such that  $\|\Delta\|_{2,2} \leq \Omega$  for all  $\Delta \in \mathcal{Z}$ .
- (B)  $\bar{A}^T \bar{A} + 2\Delta^T \bar{A}$  is positive semi-definite for all  $\Delta \in \mathcal{Z}$ .

In addition to the discussion in Appendix C on how we can check these assumptions for a general uncertainty set  $\mathcal{Z}$ , here we show how to check them on two typically used uncertainty sets.

**Ellipsoidal uncertainty set:** Let us assume that  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_{2,2} \leq \rho\}$ , for some  $\rho > 0$ . Clearly, Assumption (A) holds with  $\Omega = \rho$ . Furthermore, Assumption (B) can be checked using the following proposition:

**Proposition 1.** Let us assume that  $\bar{A} \in \mathbb{R}^{m \times n}$  is full column-rank and  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_{2,2} \leq \rho\}$ . Then, Assumption (B) holds if and only if  $\lambda_{\min}(\bar{A}^T \bar{A}) \geq 4\rho^2$ , where  $\lambda_{\min}(\bar{A}^T \bar{A})$  denotes the smallest eigenvalue of  $\bar{A}^T \bar{A}$ .

*Proof.* Proof. Using Theorem 8.2.3 in [8], Assumption (B) holds if and only if there exists a positive  $\lambda$  such that

$$\begin{bmatrix} \lambda I_n & \rho \bar{A} \\ \rho \bar{A}^T & \bar{A}^T \bar{A} - \lambda I_n \end{bmatrix} \succeq 0_{2n \times 2n}.$$

Using the Schur Complement lemma (see, e.g., Appendix A.5.5 in [16]) the above linear matrix inequality is equivalent to  $(\lambda - \rho^2) \bar{A}^T \bar{A} \succeq \lambda^2 I_n$ . Hence, Assumption (B) holds if and only if

$$\begin{aligned} & \exists \lambda > \rho^2 : \quad \bar{A}^T \bar{A} \succeq \frac{\lambda^2}{\lambda - \rho^2} I_n \\ \Leftrightarrow & \exists \lambda > \rho^2 : \quad \lambda_{\min}(\bar{A}^T \bar{A}) \geq \frac{\lambda^2}{\lambda - \rho^2} \\ \Leftrightarrow & \lambda_{\min}(\bar{A}^T \bar{A}) \geq 4\rho^2, \end{aligned}$$

where  $\lambda_{\min}(\bar{A}^T \bar{A})$  denotes the smallest eigenvalue of  $\bar{A}^T \bar{A}$ , and where the last equivalence holds since  $\frac{\lambda^2}{\lambda - \rho^2}$  is convex in  $\lambda$  (for  $\lambda > \rho^2$ ) with the minimum value of  $4\rho^2$ .  $\square$

**Box uncertainty set:** Let us assume that  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_{\infty} \leq \rho\}$ , for some  $\rho > 0$ . Using the following proposition, one can see that Assumption (A) holds with  $\Omega = \rho\sqrt{nm}$ .

**Proposition 2.** *Let  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_{\infty} \leq \rho\}$ . Then,  $\sup_{\Delta \in \mathcal{Z}} \|\Delta\|_{2,2} = \rho\sqrt{nm}$ .*

*Proof.* Proof. This follows directly from the definition of  $\ell_{\infty}$  and spectral norms.  $\square$

Moreover, using Proposition 9 in Appendix C, Assumption (B) holds if

$$\min_{y \in \mathbb{R}^n} \{y^T \bar{A}^T \bar{A} y - 2\rho \mathbb{1}^T \bar{A} y : \bar{A} y \geq 0, \|y\|_1 \leq 1\} \geq 0, \quad (19)$$

where  $\mathbb{1} \in \mathbb{R}^m$  is a vector whose components are all one.

**Remark 2.** *Notice that if the uncertainty set is  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_1 \leq \rho\}$ , for some  $\rho > 0$ , then (5) can be reformulated to a system of  $2mn$  deterministic (conic) quadratic constraints because the uncertainty set contains  $2mn$  vertices.*

Now, we proceed to the main results of this section. The following theorem provides tractable inner approximations of the constraints in the forms (5) by replacing the quadratic term in the uncertain parameter with a linear upper bound.

**Theorem 2.** *Let Assumption (A) hold. Then:*

(I)  $y \in \mathbb{R}^n$  satisfies (5a) if there exists  $W \in \mathbb{R}^{n \times n}$  satisfying the convex system

$$\text{trace}((\bar{A}^T \bar{A} + \Omega^2 I_n)W) + \delta_{\mathcal{Z}}^*(2\bar{A}W + D^T y a^T) + \bar{b}^T y + c \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad (20)$$

(II)  $y \in \mathbb{R}^n$  satisfies (5b) if there exist  $W \in \mathbb{R}^{n \times n}$  and  $\eta \in \mathbb{R}$  satisfying the convex system

$$\text{trace}((\bar{A}^T \bar{A} + \Omega^2 I_n)W) + \delta_{\mathcal{Z}}^*(2\bar{A}W + D^T y a^T) + \bar{b}^T y + c + \frac{\eta}{4} \leq 0, \quad \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad (21)$$

*Proof.* Proof. (I)  $y \in \mathbb{R}^n$  satisfies (5a) if and only if

$$y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + \|\Delta y\|_2^2 + (D\Delta a)^T y + \bar{b}^T y + c \leq 0, \quad \forall \Delta \in \mathcal{Z}. \quad (22)$$

Replacing  $\|\Delta y\|_2^2$  by its upper bound  $\Omega^2 \|y\|_2^2$  implies that  $y \in \mathbb{R}^n$  satisfies (22) if it satisfies

$$y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + \Omega^2 y^T y + (D\Delta a)^T y + \bar{b}^T y + c \leq 0, \quad \forall \Delta \in \mathcal{Z}. \quad (23)$$

Setting  $\mathcal{U} = \{(\bar{A}^T \bar{A} + 2\bar{A}^T \Delta + \Omega^2 I_n, D\Delta a) : \Delta \in \mathcal{Z}\}$ , (23) is equivalent to

$$y^T B y + (\bar{b} + d)^T y + c \leq 0 \quad \forall (B, d) \in \mathcal{U}.$$

For any  $(B, d) \in \mathcal{U}$ ,  $B$  is positive semi-definite since  $B = (\bar{A} + \Delta)^T(\bar{A} + \Delta) + \Omega^2 I_n - \Delta^T \Delta \succeq 0_{n \times n}$ . So, by applying Theorem 1(I) and Lemma 2(iii),  $y \in \mathbb{R}^n$  satisfies (22) if there exists  $W \in \mathbb{R}^{n \times n}$  such that  $y$  and  $W$  satisfy (20).

(II) The proof is similar to part (I).  $\square$

In the next theorem we derive tractable outer approximations of the constraints in the forms (5).

**Theorem 3.** *Let Assumption (B) holds. Then:*

(I) *if  $y$  satisfies (5a), then there exists  $W \in \mathbb{R}^{n \times n}$  satisfying the convex system*

$$\text{trace}(\bar{A}^T \bar{A} W) + \delta_{\mathcal{Z}}^*(2\bar{A} W + D^T y a^T) + \bar{b}^T y + c \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad (24)$$

(II) *if  $y \in \mathbb{R}^n$  satisfies (5b), then there exist  $W \in \mathbb{R}^{n \times n}$  and  $\eta \in \mathbb{R}$  satisfying the convex system*

$$\text{trace}(\bar{A}^T \bar{A} W) + \delta_{\mathcal{Z}}^*(2\bar{A} W + D^T y a^T) + \bar{b}^T y + c + \frac{\eta}{4} \leq 0, \quad \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1}. \quad (25)$$

*Proof.* Proof. (I) It is clear that  $y$  satisfies (5a) if and only if  $y$  satisfies (22). Replacing  $\|\Delta y\|_2^2$  with its lower bound 0 implies that if  $y \in \mathbb{R}^n$  satisfies (22) then

$$y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + (D\Delta a)^T y + \bar{b}^T y + c \leq 0, \quad \forall \Delta \in \mathcal{Z}. \quad (26)$$

Setting  $\mathcal{U} = \{(\bar{A}^T \bar{A} + 2\bar{A}^T \Delta, D\Delta a) : \Delta \in \mathcal{Z}\}$ , and using Theorem 1(I) and Lemma 2(iii) completes the proof.

(II) The proof is similar to the previous part.  $\square$

In the next theorem we provide an upper bound on the violation errors of (5a) and (5b) for the solutions that satisfy the outer approximations (24) and (25), respectively.

**Theorem 4.** *Let Assumptions (A) and (B) hold. Then,*

(I) *if  $y \in \mathbb{R}^n$  and  $W \in \mathbb{R}^{n \times n}$  satisfy (24), then  $y$  violates (5a) by at most  $\Omega^2 \|y\|_2^2$ .*

(II) *if  $y \in \mathbb{R}^n$  and  $W \in \mathbb{R}^{n \times n}$  satisfy (25), then  $y$  violates (5b) by at most  $\Omega \|y\|_2$ .*

*Proof.* Proof. (I) Let  $y \in \mathbb{R}^n$  and  $W \in \mathbb{R}^{n \times n}$  satisfy (24). Then,  $y$  satisfies (26). Therefore,

$$\max_{\Delta \in \mathcal{Z}} \{y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + (D\Delta a)^T y + \bar{b}^T y + c\} \leq 0. \quad (27)$$

As it is mentioned in the proof of Theorem 2(I), (5a) is equivalent to (22). Therefore, we have

$$\begin{aligned} & \max_{\Delta \in \mathcal{Z}} \{y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + \|\Delta y\|_2^2 + (D\Delta a)^T y + \bar{b}^T y + c\} \\ & \leq y^T \bar{A}^T \bar{A} y + \bar{b}^T y + c + \max_{\Delta \in \mathcal{Z}} \{2y^T \bar{A}^T \Delta y + (D\Delta a)^T y\} + \max_{\Delta \in \mathcal{Z}} \|\Delta y\|_2^2 \\ & \leq y^T \bar{A}^T \bar{A} y + \bar{b}^T y + c + \max_{\Delta \in \mathcal{Z}} \{2y^T \bar{A}^T \Delta y + (D\Delta a)^T y\} + \Omega^2 \|y\|_2^2 \leq \Omega^2 \|y\|_2^2, \end{aligned}$$

where the last inequality follows from (27).

(II) It is clear that (5b) is equivalent to

$$\sqrt{y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + \|\Delta y\|_2^2 + (D\Delta a)^T y + \bar{b}^T y + c} \leq 0, \quad \forall \Delta \in \mathcal{Z}.$$

Similar to the previous part, if  $y$  comes from the outer approximation (25), then we have

$$\begin{aligned} & \sqrt{\max_{\Delta \in \mathcal{Z}} \{y^T \bar{A}^T \bar{A} y + 2y^T \bar{A}^T \Delta y + \|\Delta y\|_2^2 + (D\Delta a)^T y\} + \bar{b}^T y + c} \\ & \leq \sqrt{y^T \bar{A}^T \bar{A} y + \max_{\Delta \in \mathcal{Z}} \{2y^T \bar{A}^T \Delta y + (D\Delta a)^T y\}} + \max_{\Delta \in \mathcal{Z}} \|\Delta y\|_2 + \bar{b}^T y + c \\ & \leq \sqrt{y^T \bar{A}^T \bar{A} y + \max_{\Delta \in \mathcal{Z}} \{2y^T \bar{A}^T \Delta y + (D\Delta a)^T y\}} + \Omega \|y\|_2 + \bar{b}^T y + c \leq \Omega \|y\|_2, \end{aligned}$$

where the first inequality holds because of the fact that  $\sqrt{f+g} \leq \sqrt{f} + \sqrt{g}$  for any  $f, g \geq 0$ .  $\square$

**Remark 3.** *Until now, we have considered problems containing uncertainties in their constraint parameters. This is without loss of generality, since if we have a problem with uncertainty in the parameters of the objective function, then we can use the epigraph formulation to shift the uncertainty to a constraint.*  $\square$

## 6 Data-driven uncertainty set

A usual way of constructing an uncertainty set is by using historical data and statistical tools, such as hypothesis testing ([11]), or asymptotic confidence sets ([7]). In this section, we use the latter to design an uncertainty set for a vector consisting of the mean and vectorized covariance matrix.

For notational simplicity, we explain how to construct an uncertainty set for the two dimensional case; the extension to higher dimensions is straightforward. For the two dimensional case, assume that  $\begin{pmatrix} x \\ z \end{pmatrix}$  is a random vector with components  $x, z$  and set  $\mu_x = \mathbb{E}(x)$ ,  $\mu_z = \mathbb{E}(z)$ ,  $\sigma_x^2 = \mathbb{E}(x - \mu_x)^2$ ,  $\sigma_z^2 = \mathbb{E}(z - \mu_z)^2$ ,  $\sigma_{xz} = \mathbb{E}(x - \mu_x)(z - \mu_z)$ , and  $\mu_{kl} = \mathbb{E}(x - \mu_x)^k(z - \mu_z)^l$ ,  $k, l = 0, 1, 2, \dots$ . Assume that the fourth moments exist, which means that  $\mu_{kl}$  exists when  $k + l \leq 4$ ,  $k, l = 0, 1, 2, 3, 4$ . This assumption can be tested using the result in [38]. Now, consider a random sample of size  $n$ ,  $\begin{pmatrix} x_i \\ z_i \end{pmatrix}$ ,  $i = 1, \dots, n$ . Set

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{z} = \frac{1}{n} \sum_{i=1}^n z_i, \quad S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad S_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2, \quad S_{xz} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z}).$$

Using the Central Limit Theorem (Example 2.18 in [39]) and the Delta Method (Theorem 3.1 in [39]), and setting

$$\mathcal{Y} = (\mu_x, \mu_z, \mathbb{E}(x^2), \mathbb{E}(xz), \mathbb{E}(z^2))^T, \quad Y_n = \left( \bar{x}, \bar{z}, \frac{1}{n} \sum_{i=1}^n x_i^2, \frac{1}{n} \sum_{i=1}^n x_i z_i, \frac{1}{n} \sum_{i=1}^n z_i^2 \right)^T,$$

it follows for any differentiable function  $\phi: \mathbb{R}^5 \rightarrow \mathbb{R}^m$  that  $\sqrt{n}(\phi(Y_n) - \phi(\mathcal{Y}))$  converges in distribution to the normal distribution  $N(0, \nabla\phi(\theta)\Sigma\nabla\phi(\theta)^T)$ , where  $\Sigma$  and  $\nabla\phi$  are the covariance matrix of  $(x, z, x^2, xz, z^2)^T - \mathcal{Y}$  and the Jacobian matrix of  $\phi$ , respectively. Letting

$$\phi(x_1, \dots, x_5) = (x_1, x_2, x_3 - x_1^2, x_4 - x_1 x_2, x_5 - x_2^2)^T,$$

it is easy to show, similar to Example 3.2 in [39], that

$$\sqrt{n}(T_n - \theta) \xrightarrow[n \rightarrow \infty]{d} N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \underbrace{\begin{pmatrix} \mu_{20} & \mu_{11} & \mu_{30} & \mu_{21} & \mu_{12} \\ \mu_{11} & \mu_{02} & \mu_{21} & \mu_{12} & \mu_{03} \\ \mu_{30} & \mu_{21} & \mu_{40} - \mu_{20}^2 & \mu_{31} - \mu_{11}\mu_{20} & \mu_{22} - \mu_{20}\mu_{02} \\ \mu_{21} & \mu_{12} & \mu_{31} - \mu_{11}\mu_{20} & \mu_{22} - \mu_{11}^2 & \mu_{13} - \mu_{11}\mu_{02} \\ \mu_{12} & \mu_{03} & \mu_{22} - \mu_{20}\mu_{02} & \mu_{13} - \mu_{11}\mu_{02} & \mu_{04} - \mu_{02}^2 \end{pmatrix}}_V \right), \quad (28)$$

where  $\theta = \phi(\mathcal{Y}) = (\mu_x, \mu_z, \sigma_x^2, \sigma_{xz}, \sigma_z^2)^T$ ,  $T_n = \phi(\mathcal{Y}_n) = (\bar{x}, \bar{z}, S_x^2, S_{xz}, S_z^2)^T$ , and  $\xrightarrow[n \rightarrow \infty]{d}$  means convergence in distribution when the size of the random sample goes to infinity.

Let  $\hat{V}$  and  $\hat{\theta}$  be consistent estimates of  $V$  and  $\theta$  defined in (28), respectively. Then, asymptotically with  $(1 - \alpha)\%$  confidence,  $\theta$  belongs to the following ellipsoid:

$$\mathcal{U} := \left\{ \theta : n(\hat{\theta} - \theta)^T \hat{V}^{-1} (\hat{\theta} - \theta) \leq \chi_{rank(V), 1-\alpha}^2 \right\},$$

where  $\chi_{d, 1-\alpha}^2$  denotes the  $(1 - \alpha)$  percentile of the Chi-square distribution with  $d$  degrees of freedom.

To use the results of Section 4, we reformulate the uncertainty set  $\mathcal{U}$ . Setting

$$\Psi = \begin{bmatrix} \Psi_\mu \\ \Psi_\Sigma \end{bmatrix}, \quad \Psi_\mu = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad \Psi_\Sigma = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mu = [\mu_x \ \mu_z]^T, \quad \Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xz} \\ \sigma_{xz} & \sigma_z^2 \end{bmatrix}, \quad (29)$$

due to positive semi-definiteness of  $\Sigma$ , with  $(1 - \alpha)\%$  confidence

$$\begin{pmatrix} \mu \\ svec(\Sigma) \end{pmatrix} \in \bar{\mathcal{U}} := \hat{\mathcal{U}} \cap \left\{ \gamma : n \left( \Psi \hat{\theta} - \gamma \right)^T \Psi^{-1} \hat{V}^{-1} \Psi^{-1} \left( \Psi \hat{\theta} - \gamma \right) \leq \chi_{rank(V), 1-\alpha}^2 \right\},$$

where  $\hat{\mathcal{U}} = \left\{ \begin{pmatrix} \gamma_\mu \\ \gamma_\Sigma \end{pmatrix} : \gamma_\Sigma = svec(M), M \succeq 0_{n \times n} \right\}$ . Letting  $R^T R$  be the Cholesky factorization of  $\hat{V}^{-1}$ , i.e.,  $\hat{V}^{-1} = R^T R$ ,  $\bar{\mathcal{U}}$  can be rewritten as

$$\begin{aligned} \bar{\mathcal{U}} &= \hat{\mathcal{U}} \cap \left\{ \gamma : \left\| R \Psi^{-1} \left( \gamma - \Psi \hat{\theta} \right) \right\|_2 \leq \sqrt{\frac{\chi_{rank(V), 1-\alpha}^2}{n}} \right\} \\ &= \hat{\mathcal{U}} \cap \left\{ \Psi R^{-1} \nu + \Psi \hat{\theta} : \|\nu\|_2 \leq \sqrt{\frac{\chi_{rank(V), 1-\alpha}^2}{n}} \right\}. \end{aligned}$$

Hence, by letting the estimated mean vector and covariance matrix based on the random sample be  $\hat{\mu}$  and  $\hat{\Sigma}$ , respectively, we have

$$\bar{\mathcal{U}} = \hat{\mathcal{U}} \cap \left\{ \Psi R^{-1} \nu + \begin{pmatrix} \hat{\mu} \\ svec(\hat{\Sigma}) \end{pmatrix} : \|\nu\|_2 \leq \sqrt{\frac{\chi_{rank(V), 1-\alpha}^2}{n}} \right\}. \quad (30)$$

**Remark 4.** If  $V$  is not invertible, then one can use a generalized inverse, such as the Moore-Penrose inverse.  $\square$

**Remark 5.** The construction of the uncertainty set can straightforwardly be extended to higher dimensions using suitable  $\phi$ ,  $\Psi$ , and  $V$ . Details are omitted for brevity of exposition.  $\square$

**Remark 6.** The uncertainty set  $\bar{\mathcal{U}}$  is constructed for a random sample. Analogously, one can construct an uncertainty set for a time-series under appropriate assumptions; see, e.g., Section 2.2 in the book [30].  $\square$

Now, consider a convex quadratic constraint

$$y^T \Sigma y + \mu^T y + c \leq 0, \quad (31)$$

where  $\mu$  and  $\Sigma$  are the mean vector and covariance matrix of a random vector. By using the uncertainty set  $\bar{\mathcal{U}}$  in (30) and Example 2, the RC of (31) is

$$\hat{\mu}^T y + trace \left( \hat{\Sigma} W \right) + \rho \left\| \left( \Psi R^{-1} \right)^T \begin{pmatrix} y \\ svec(W) \end{pmatrix} \right\|_2 + c \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}, \quad (32)$$

where  $\rho = \sqrt{\frac{\chi_{rank(V), 1-\alpha}^2}{n}}$ .

Let  $\mu$  and  $\Sigma$  be the actual population mean vector and covariance matrix, respectively. Then,  $\theta = \begin{pmatrix} \mu \\ svec(\Sigma) \end{pmatrix}$  belongs to the uncertainty set  $\bar{\mathcal{U}}$  asymptotically with confidence level  $(1 - \alpha)\%$ . This, roughly speaking, means that the uncertainty set not only contains  $\theta$  but also many more points. Therefore,  $y$  that satisfies (32) is asymptotically immunized against some extra  $\mu$  and  $\Sigma$  and hence conservative.

Another way of dealing with the uncertainty in  $\theta$  is by making use of the chance constraint  $Prob(y^T \Sigma y + \mu^T y + c \leq 0) \geq 1 - \alpha$ , where  $\alpha > 0$  is close to 0. In what follows, we elaborate more on this chance constraint and provide a reformulation and relaxation of it.

For any vector  $\beta$ , (28) implies that  $\sqrt{n} (\beta^T T_n - \beta^T \theta) \xrightarrow[n \rightarrow \infty]{d} N(0, \beta^T V \beta)$ . By setting  $\beta = \Psi \begin{pmatrix} y \\ svec(y y^T) \end{pmatrix}$ , it follows straightforwardly that the (asymptotic) chance constraint with probability of  $1 - \alpha$  is equivalent to

$$\frac{z_{1-\alpha}}{\sqrt{n}} \sqrt{\beta^T \hat{V} \beta} + \hat{\mu}^T y + y^T \hat{\Sigma} y + c \leq 0, \quad (33)$$

where  $z_{1-\alpha}$  is the  $1 - \alpha$  percentile of the standard normal distribution. Clearly (33) is equivalent to the set of constraints

$$\frac{z_{1-\alpha}}{\sqrt{n}} \|R^{T^{-1}} \beta\| + \hat{\mu}^T y + y^T \hat{\Sigma} y + c \leq 0, \quad \beta = \Psi \begin{pmatrix} y \\ \text{svec}(W) \end{pmatrix}, \quad W = yy^T,$$

where  $R$  is the Cholesky factorization of  $\hat{V}^{-1}$ . The constraint  $W = yy^T$  is nonconvex, so we relax it to  $W \succeq yy^T$ , which is a semi-definite representable constraint. Hence,

$$\frac{z_{1-\alpha}}{\sqrt{n}} \|R^{T^{-1}} \beta\| + \hat{\mu}^T y + y^T \hat{\Sigma} y + c \leq 0, \quad \beta = \Psi \begin{pmatrix} y \\ \text{svec}(W) \end{pmatrix}, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}, \quad (34)$$

is a relaxation of (33). In the next proposition, we provide a relation between solutions that satisfy (32) and the ones satisfying (34).

**Proposition 3.** *Let  $(y, W)$  be a solution that satisfies (32). Then  $(y, W)$  also satisfies (34).*

*Proof.* Proof. Appendix A.5. □

Even though Proposition 3 asserts that (32) is more conservative than (34), we cannot conclude that (32) is more conservative than (33). This is because  $W = yy^T$  is not necessarily satisfied for solutions of (34).

**Remark 7.** *After solving the problem containing (34), if  $W = yy^T$  is not satisfied, then  $y$  is not feasible for (33). However,  $y$  is strictly feasible for*

$$y^T \Sigma y + \mu^T y + c \leq 0 \quad (35)$$

*for the nominal scenario  $(\hat{\mu}, \hat{\Sigma})$ . Therefore,  $y$  may also be feasible for other scenarios and hence is more robust than the nominal solution.* □

Consider a solution  $\bar{y}$  that satisfies (32) where the uncertainty set is constructed using the desired confidence level  $1 - \bar{\alpha}$ . For this solution, according to the above discussion,  $\text{Prob}(\bar{y}^T \Sigma \bar{y} + \mu^T \bar{y} + c \leq 0)$  might be larger than the desired confidence level  $1 - \bar{\alpha}$ . If so, then by decreasing the confidence level that is used in the construction of the uncertainty set and considering  $\tilde{y}$  that satisfies (32),  $\text{Prob}(\tilde{y}^T \Sigma \tilde{y} + \mu^T \tilde{y} + c \leq 0)$  gets closer to the desired confidence level  $1 - \bar{\alpha}$ . In our numerical experiments, we check for different instances which confidence level should be used in the construction of the uncertainty set such that for the robust solution the constraint  $y^T \Sigma y + \mu^T y + c \leq 0$  is satisfied with probability close to the desired confidence level.

## 7 Applications

In this section, we apply the results of the previous sections to a robust portfolio choice, norm approximation, and regression line problem. All computations in this paper were carried out with MATLAB 2016a using YALMIP [32] to pass the optimization problems to MOSEK 8.1.0.80 [33].

### 7.1 Mean-Variance portfolio problem

In this subsection, we describe a formulation for a mean-variance portfolio problem (Chapter 2 in [21]), and use the results of Section 6 to construct an uncertainty set and to derive a tractable reformulation of the robust counterpart.

**Problem formulation:** We consider a mean-variance portfolio problem with  $n$  assets. Let  $\mu$  and  $\Sigma$  be the expectation and covariance matrix of the return vector  $r = (r_1, \dots, r_n)$ , respectively. One formulation of a mean-variance portfolio problem is to model the trade-off between the risk and mean return in the objective function using a risk-aversion coefficient  $\lambda$ :

$$\max_{\omega} \{ \mu^T \omega - \lambda \omega^T \Sigma \omega : \mathbf{1}^T \omega = 1, \omega \geq 0 \}, \quad (36)$$

where  $\mathbf{1} = [1, 1, \dots, 1]^T$ . The risk aversion coefficient is determined by the decision maker. When it is small, it means that the mean return is more important than the corresponding risk and it leads to a more risky portfolio than when the risk-aversion coefficient is large.

In practice  $\mu$  and  $\Sigma$  are typically estimated from a set of historical data, which makes them sensitive to sampling inaccuracy. There are several ways of defining uncertainty sets for the expected return vector and asset return covariance matrix, e.g., see Chapter 12 in [21]. In this section, we use  $\bar{U}$  defined in (30), i.e., the uncertainty set constructed for  $\begin{pmatrix} \mu \\ svec(\Sigma) \end{pmatrix}$ . Using (32), the robust counterpart of (36) with uncertainty set  $\bar{U}$  reads

$$\begin{aligned} \max_{\omega, W} \quad & \hat{\mu}^T \omega - \lambda tr(\hat{\Sigma}W) - \rho \left\| (\Psi R^{-1})^T \begin{pmatrix} -\omega \\ \lambda svec(W) \end{pmatrix} \right\|_2 \\ \text{s.t.} \quad & \begin{bmatrix} W & \omega \\ \omega^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}, \quad \mathbf{1}^T \omega = 1, \quad \omega \geq 0, \end{aligned} \quad (37)$$

where  $\rho = \sqrt{\frac{\chi_{rank(V), 1-\alpha}^2}{n}}$ ,  $\hat{\mu}$ ,  $\Sigma$ , and  $\hat{V}$  are consistent estimates of  $\mu$ ,  $\Sigma$ , and  $V$ , with  $V$  and  $\Psi$  as in (28) and (29), respectively, but formulated for the higher dimensional case, and  $R$  is the Cholesky factorization of  $\hat{V}^{-1}$ .

Furthermore, by setting  $\beta = \Psi \begin{pmatrix} \omega \\ -\lambda svec(W) \end{pmatrix}$ , and using the relaxed chance constraint (34), the robust counterpart of problem (36) with confidence  $(1 - \alpha)\%$  is approximated by

$$\begin{aligned} \max_{\omega} \quad & \hat{\mu}^T \omega - \lambda \omega^T \hat{\Sigma} \omega - \frac{z_{1-\alpha}}{\sqrt{n}} \left\| R^{T^{-1}} \beta \right\|_2 \\ \text{s.t.} \quad & \begin{bmatrix} W & \omega \\ \omega^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}, \quad \mathbf{1}^T \omega = 1, \quad \omega \geq 0. \end{aligned} \quad (38)$$

**Numerical evaluation:** To evaluate the above robust counterparts, we use the monthly average value weighted return of 5 and 30 industries from 1956 until 2015, obtained from ‘‘Industry Portfolios’’ data on the website [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The data are monthly returns, but to present the results, we report the annualized returns (obtained by multiplying the expected monthly return by 12) and the annualized risk (multiplication of the standard deviation by  $\sqrt{12}$ ). Furthermore, we set the risk aversion coefficient  $\lambda$  to 3.

We have solved the following three problems: (36) with nominal values for  $\mu$  and  $\Sigma$  estimated from the data, which we call *Nominal problem*; (37), which we call *Robust problem*; and (38), which we call *Chance problem*, due to the chance constraint.

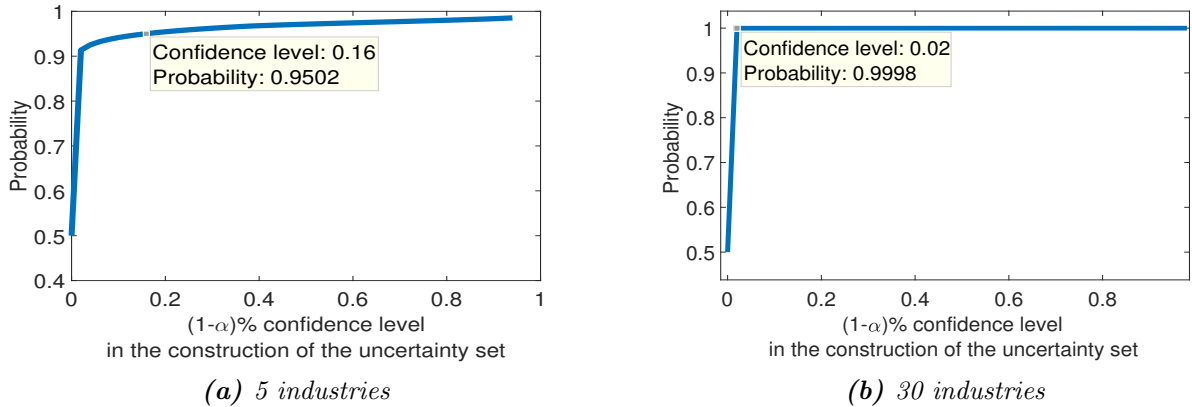
We first check the behavior of  $Prob(\mu^T \omega^* - \lambda \omega^{*T} \Sigma \omega^* \geq z^*)$  as a function of the confidence level used to construct the uncertainty set, where  $\omega^*$  and  $z^*$  are the robust solution and corresponding robust objective value, respectively. As shown in Figure 1, in order to be sure that the constraint  $\mu^T \omega^* - \lambda \omega^{*T} \Sigma \omega^* \geq z^*$  is satisfied with probability of at least 95%, one can reduce the confidence level used in the construction of the uncertainty set from 95% to 16% for the 5 industries case, and to 2% for the 30 industries case.

We emphasize that even though the confidence level of 2% seems to make the uncertainty set much smaller than the one corresponding to 95%, this does not happen for the 30 industries case since we have  $\sqrt{\frac{\chi_{rank(V), 0.95}^2}{n}} = 0.8723$  and  $\sqrt{\frac{\chi_{rank(V), 0.02}^2}{n}} = 0.7751$ , where  $n = 720$  and  $rank(V) = 495$ .

**Remark 8.** *The 2% confidence level was achieved from Figure 1(b), which was plotted by discretizing  $[0, 1]$  into the set of points starting from 0 with the step size of 0.02. Hence, choosing a smaller step size may result in a smaller confidence level. However, the uncertainty set will not be much smaller than the one constructed by 2% confidence level, as it can be easily checked that even for 10<sup>-23</sup>% confidence level we have  $\sqrt{\frac{\chi_{rank(V), 10^{-25}}^2}{n}} = 0.5710$ , where  $n = 720$  and  $rank(V) = 495$ .*  $\square$

We considered both data sets with 5 and 30 industries in our numerical experiments; however, due to similarity in the results, we present the results of considering only data set with 30 industries.





**Figure 1.** The horizontal axis presents the value of  $(1 - \alpha)\%$ , the confidence level used in the uncertainty set for data with 5 and 30 industries. The vertical axis presents the value of  $\text{Prob}(\mu^T \omega^* - 3\omega^{*T} \Sigma \omega^* \geq z^*)$ , where  $\omega^*, z^*$  are the robust solution and corresponding objective value for (37), respectively, with the uncertainty set (30) and different  $\alpha$ . The plots are constructed by considering multiplications of 0.02 in  $[0, 1]$  as values of  $\alpha$ .

After solving the *Nominal problem*, the *Robust problem* considering the uncertainty set with 95% confidence level, the *Robust problem* considering the uncertainty set with 2% confidence level, and the *Chance problem* with 95% confidence level, we compare the solutions in three ways:

- (i) evaluating the solutions with respect to the nominal values;
- (ii) evaluating the solutions with respect to their worst-case scenarios in the uncertainty set constructed with 95% confidence level;
- (iii) evaluating the solutions with respect to their worst-case scenarios in the uncertainty set constructed with 2% confidence level.

Table 1 presents the evaluations of the solutions. In the first block row (with results), the evaluation is done using the nominal scenario. The objective value of the *Nominal problem* is the highest. The worst objective value in this row is corresponding to the solution of the *Robust problem* considering the uncertainty set with 95% confidence level. This solution is immunized against more scenarios than the others.

The second block row is the evaluation of the solutions considering their worst-case scenario in the uncertainty set constructed by 95% confidence level. This implies that the solution of the *Robust problem* with this uncertainty set has the highest objective value, because the solution is immunized against all scenarios in the uncertainty set; however, other solutions are immunized against all scenarios in a subset of the uncertainty set. The third block row has the same interpretation, where the scenario is chosen in the uncertainty set with confidence level 2%.

Table 1 shows that even though all solutions have close annualized returns and risks in the nominal scenario, the solutions of (37) have extremely better returns and risks in the included worst-case scenarios.

Proposition 3 states that a solution of (37), denoted by  $(\bar{\omega}, \bar{W})$ , is more conservative than a solution of (38), denoted by  $(\tilde{\omega}, \tilde{W})$ . This means  $(\bar{\omega}, \bar{W})$  is safeguarded against more scenarios (all scenarios in the uncertainty set  $\bar{U}$ ) than  $(\tilde{\omega}, \tilde{W})$ . Therefore, as the last column of Table 1 shows, the objective values of  $(\bar{\omega}, \bar{W})$  at their worst-case scenarios in  $\bar{U}$  are worse than the ones for  $(\tilde{\omega}, \tilde{W})$ .

## 7.2 Least-squares problems with uncertainties

This subsection contains applications of the results of Section 5 to two well-known problems, namely a norm approximation and a linear regression problem.

		solution of	solution of		solution of
		<i>Nominal problem</i> (36)	<i>Robust problem (37)</i> with confidence level		<i>Chance problem</i> (38)
			95%	2%	
Nominal case	Obj. value	<b>-35.57</b>	-39.24	-38.84	-35.59
	<i>Ann. risk</i>	<i>12.03</i>	<i>12.63</i>	<i>12.56</i>	<i>12.03</i>
	<i>Ann. return</i>	<i>7.02</i>	<i>7.29</i>	<i>7.28</i>	<i>6.98</i>
Worst-case with confidence level 95%	Obj. value	-77.47	<b>-51.11</b>	-51.12	-55.01
	Ann. risk	<i>15.77</i>	<i>13.22</i>	<i>13.20</i>	<i>13.42</i>
	Ann. return	<i>-183.10</i>	<i>-89.14</i>	<i>-90.49</i>	<i>-119.87</i>
Worst-case with confidence level 2%	Obj. value	-75.00	-50.39	<b>-50.38</b>	-53.89
	<i>Ann. risk</i>	<i>15.57</i>	<i>13.18</i>	<i>13.16</i>	<i>13.33</i>
	<i>Ann. return</i>	<i>-172.72</i>	<i>-83.91</i>	<i>-85.20</i>	<i>-113.37</i>

**Table 1.** Comparison among the solutions of the nominal problem (36), the Robust problem (37) considering the uncertainty set with 95% confidence level, the Robust problem (37) considering the uncertainty set with 2% confidence level, and (38) in three way: The first block row with results is the nominal evaluation of the solutions. The second and third block rows are the evaluation of the solutions with respect to their worst-case scenarios in uncertainty sets 95%, and 2% confidence level, respectively. The results are by considering the data for 30 industries. The **bold numbers** shows the best objective value in each scenario. The annualized return and risk are in italics and not individually optimized.

### 7.2.1 Norm approximation with uncertainty in the coefficients

The norm approximation  $\min_{y \in \mathbb{R}^n} \|Ay - b\|_2$  tries to find the closest vector to  $b \in \mathbb{R}^m$  in the range of the linear function  $Ay$ . The solution to this problem can be sensitive even to small errors in  $A$  or  $b$ . To detect this, one can analyze the condition number of the matrix  $A$  and check the sensitivity of the nominal solution to a perturbation in  $A$ , see, e.g., Chapter 7 in [27]. If the condition number is large, then the solution might be sensitive to a small error in  $A$  or  $b$ , hence not reliable. In this subsection we are using the results of Section 5 to deal with this problem.

Consider the uncertain norm approximation  $\min_y \|(\bar{A} + \Delta)y - b\|_2$ , where  $\Delta \in \mathcal{Z} \subseteq \mathbb{R}^{m \times n}$  reflects the uncertainty in  $\bar{A}$ . This problem is equivalent to  $\min_{y \in \mathbb{R}^n} y^T (\bar{A} + \Delta)^T (\bar{A} + \Delta) y + 2b^T (\bar{A} + \Delta) y + b^T b$ . Now using the results of Section 5, upper and lower bounds on the robust optimal value of this problem are obtained by solving

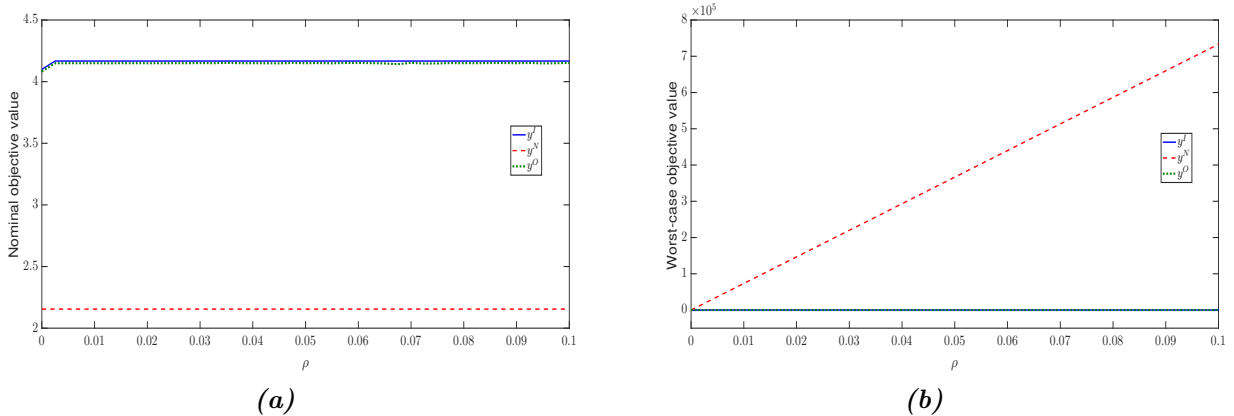
$$\min_{W, y} \left\{ \text{trace}((\bar{A}^T \bar{A} + \Omega^2 I_n)W) + \delta_{\mathcal{Z}}^*(2W\bar{A}^T - 2by^T) - 2b^T \bar{A}y + \|b\|_2^2 : \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1} \right\}, \quad (39)$$

and

$$\min_{W, y} \left\{ \text{trace}(\bar{A}^T \bar{A}W) + \delta_{\mathcal{Z}}^*(2W\bar{A}^T - 2by^T) - 2b^T \bar{A}y + \|b\|_2^2 : \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1} \right\}, \quad (40)$$

respectively.

For our numerical experiments, we construct randomly generated problems with ill-conditioned  $\bar{A}$  as follows: we fix  $n = 100$  and generate randomly a matrix  $U \in (0, 1)^{n \times n}$  and a vector  $b \in (0, 1)^n$ . Also, we randomly generate an integer  $i$  in  $\{1, \dots, n-1\}$  and construct a diagonal matrix  $D$  whose first  $i$  diagonal entries are randomly chosen in  $(-5, 5)$  and the remaining diagonal entries are randomly chosen in  $(0, 10^{-8})$ . Then, we set  $\bar{A} := U^T D U$ . Using this procedure, we generate 20 ill-conditioned  $\bar{A}$  matrices with condition numbers in the interval  $[10^{15}, 10^{18}]$ . Moreover, we generate uniformly distributed pseudorandom matrices  $B^1, B^2 \in \{0, 1\}^{n \times n}$ , and an integer number  $K \in \{1, 2, \dots, n^2\}$  using MATLAB built-in function “randi”. Then, we solve the norm approximation



**Figure 2.** The average behavior of the objective values of  $y^I$ ,  $y^O$ , and  $y^N$  related to 20 randomly generated norm approximation problems. (a) The nominal objective value is computed by  $\|\bar{A}y - b\|_2$ . (b) The worst-case objective value is computed by  $\|(\bar{A} + \Delta^*)y - b\|_2$ , where  $\Delta^*$  is the worst-case scenario corresponding to  $y^N$ ,  $y^I$  or  $y^O$ . Notice that the scales of the vertical axes in (a) and (b) are different. The solid blue, red dashed, and green dotted curves correspond to  $y^I$ ,  $y^N$ , and  $y^O$ , respectively.

problems using the generated matrices and the budget-type uncertainty set, proposed in [14]:

$$\mathcal{Z} = \{\Delta \in \mathbb{R}^{n \times n} : \|\Delta\|_\infty \leq \rho\} \cap \{\Delta \in \mathbb{R}^{n \times n} : \|B^k \circ \Delta\|_1 \leq K\rho, k = 1, 2\}, \quad (41)$$

for some  $\rho > 0$ . For this uncertainty set, one can derive  $\delta_{\mathcal{Z}}^*(U)$  using Lemma 2.(iv), 2.(vi), and 3.(a). It is worth noting that the constructed uncertain norm approximation problems contain  $100 \times 100 = 10,000$  uncertain parameters, and hence obtaining an exact optimal robust solution is computationally intractable.

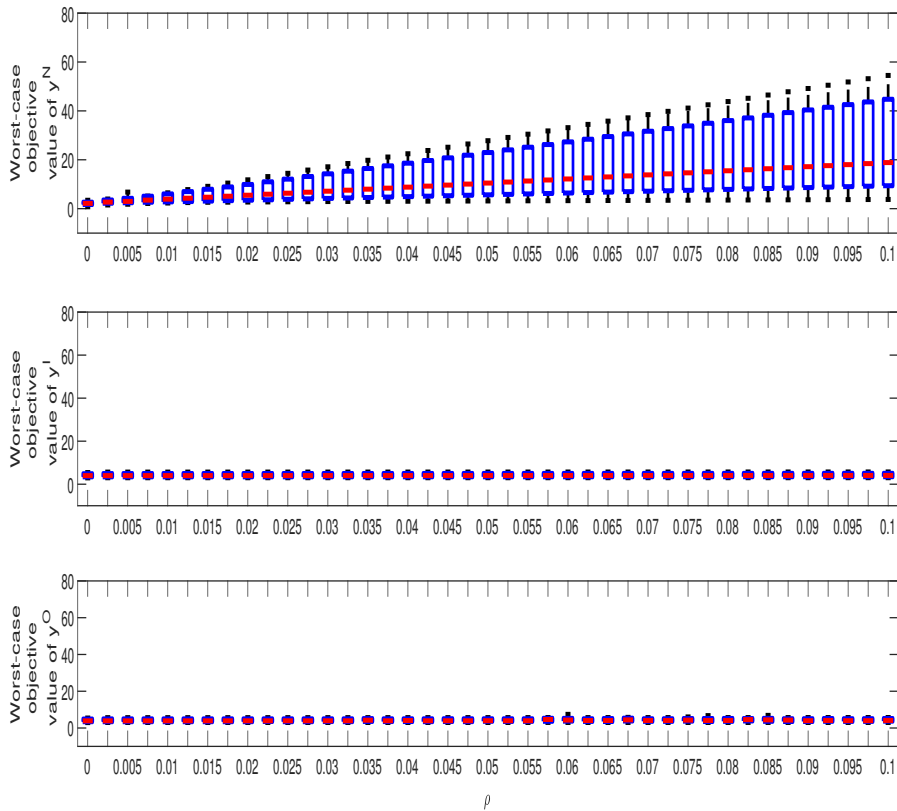
We analyze the performance of the solutions by comparing the objective values of the solutions  $y^N$ ,  $y^I$ , and  $y^O$  for both the nominal matrix  $\bar{A}$  and a worst-case matrix  $\bar{A} + \Delta^*$  corresponding to the vector  $y$ , where  $\Delta^*$  is constructed using the algorithm proposed in Appendix D.

Figure 2 provides a visualization of the average performance of  $y^N$ , the nominal solution,  $y^I$ , the solution of (39), and  $y^O$ , the solution of (40) for different scenarios and values of  $\rho \in [0, 0.1]$ . We check Assumption (B) by solving (19) and find that this assumption holds when  $\rho < 0.02$  for all instances except one. When this assumption is not satisfied,  $y^O$  is just an approximated robust solution, which is more robust than the nominal solution, and (40) is no longer a lower bound for

$$\min_{y \in \mathbb{R}^n} \max_{\Delta \in \mathcal{Z}} y^T (\bar{A} + \Delta)^T (\bar{A} + \Delta) y + 2b^T (\bar{A} + \Delta) y + b^T b.$$

One of the important observations from Figure 2 is that even though the constructed matrices are nonsingular, and hence the true nominal objective value is zero, the solver is not able to find the true optimal solution of the nominal problem because of the large condition number of the matrix  $\bar{A}$ . Furthermore, despite the small difference in the average performance of the solutions in the nominal case, the average performance of  $y^N$  in its worst-case scenario is extremely worse than the performance of  $y^I$  and  $y^O$ . For instance, for  $\alpha = 0.7$ , the average value of  $\|(\bar{A} + \Delta_N^*)y^N - b\|_2$  is  $5.13 \times 10^5$  whereas the average values of  $\|(\bar{A} + \Delta_I^*)y^I - b\|_2$  is 4.22 and the average value of  $\|(\bar{A} + \Delta_O^*)y^O - b\|_2$  is 4.29, where  $\Delta_N^*$ ,  $\Delta_I^*$ , and  $\Delta_O^*$  are the worst-case scenarios corresponding to  $y^N$ ,  $y^I$ , and  $y^O$ , respectively. This implies that in average  $(\bar{A} + \Delta_N^*)y^N$  is a point in the range of  $\bar{A} + \Delta_N^*$  that is far from  $b$ , as  $b \in (0, 1)^n$ .

Figure 3 provides the box plot of the objective values of the solutions with respect to their worst-case scenarios, where, for each  $\rho$ , the box represents the values between the first and third quartile, the dashed line above and below each box indicates the range of objective values excluding the outliers, and the red line in each box represents the second quartile.



**Figure 3.** Box plot of the worst-case objective values of  $y^N$ ,  $y^I$ , and  $y^O$  for 20 randomly generated norm approximation problems with  $\rho \in [0, 0.1]$ . The boxes are representing the values between the first and the third quartile. The outliers are left out in the figures to have a better comparison (cf. the main text where more is explained).

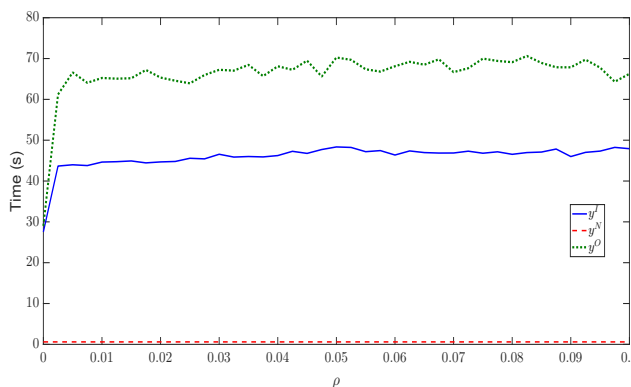
As this figure shows, the variance of the worst-case objective values of  $y^I$  and  $y^O$  does not change much as  $\rho$  increases. However, the worst-case objective value of  $y^N$  significantly changes when  $\rho$  increases. This shows the robustness of  $y^I$  and  $y^O$  against small changes in the components of  $\bar{A}$ , whereas  $y^N$  is very sensitive to these changes.

Furthermore, the comparison between Figures 2.(b) and 3 shows that the extremely high average value of  $\|(\bar{A} + \Delta_N^*)y^N - b\|_2$  is because of some outliers with extremely high values. However, even after removing the outliers,  $y^I$  and  $y^O$  outperform  $y^N$ .

Figure 4 provides the average time (in seconds) taken by MOSEK to solve the nominal problem as well as (39) and (40) to obtain  $y^N$ ,  $y^I$ , and  $y^O$ , respectively. We emphasize that even though (39) and (40) have only one constraint, which is a linear matrix inequality, we need  $O(n^2)$  more variables and constraints to pass the optimization problems to the solver. This is the reason that we see a difference between the time spent to get  $y^N$  with the one for  $y^I$  and  $y^O$  when  $\rho = 0$ .

### 7.2.2 Robust linear regression with data inaccuracy

Another application of the results of this paper is finding a robust linear regression of a dependent variable  $Y$  and a vector of independent variables  $X$  that are highly collinear. For a data set with  $n$  linearly independent variables and  $m$  data points, a mathematical formulation of finding the



**Figure 4.** Average time (in seconds) spent by MOSEK to obtain  $y^N$ ,  $y^I$ , and  $y^O$  for 20 randomly generated norm approximation problems. The solid blue, red dashed, and green dotted curves correspond to  $y^I$ ,  $y^N$ , and  $y^O$ , respectively.

regression line is

$$\min_{w,c,b} \left\{ \|w\|_2, : w_i = \sum_{j=1}^n X_{ij}c_j + b - Y_i, \forall i = 1, \dots, m \right\}, \quad (42)$$

where  $X_{ij}$  is the  $i$ -th observed value of the  $j$ -th independent variable and  $Y_i$  is the value of the dependent variable in the  $i$ -th observation.

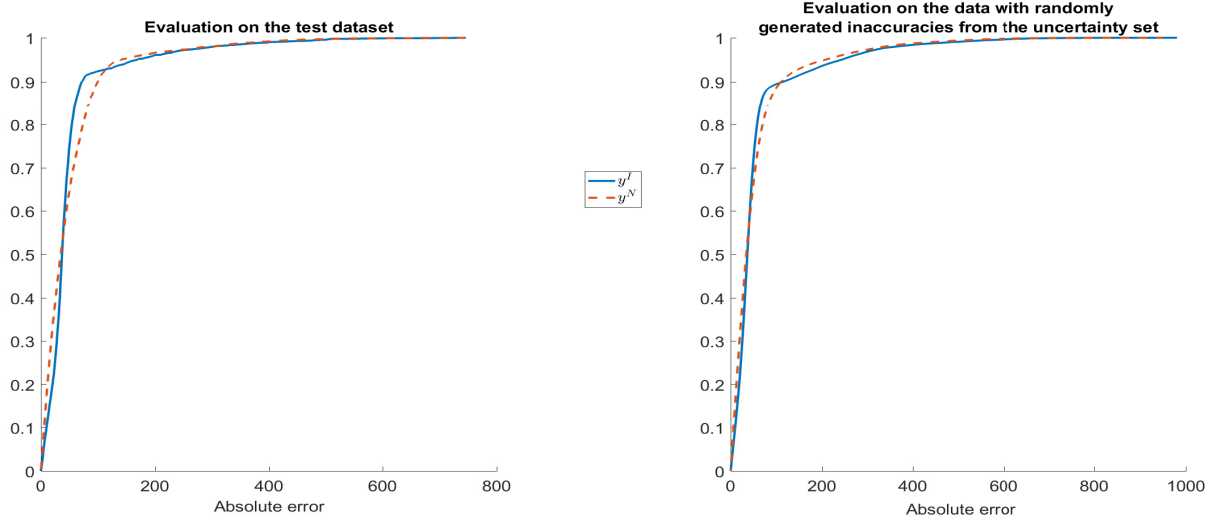
For our numerical experiment, we use the dataset proposed in [17], which is used to create a regression model of appliances energy consumption in a low-energy house located in Stambruges, Belgium. The dataset consists of 19,735 observations of 26 continuous measurable variables, 10 of which are temperatures of different parts of the house. The description of the variables can be found in Table 2 in [17]. In this section, we analyze the performance of our results in acquiring a robust linear regression model to predict the appliances energy consumption.

To reformulate (42) into the form (4b), let  $\bar{A} \in \mathbb{R}^{m \times (n+2)}$  be a matrix whose collection of the first  $n = 25$  columns is the matrix  $X$  consisting of the observations corresponding to all variables except the appliances energy consumption, the  $(n+1)$ st = 26th column corresponds to the observations of the appliances energy consumption, and the components of the last column are all ones. Then, problem (42) is equivalent to  $\min_{y \in \mathbb{R}^{(n+2)}} \{\|\bar{A}y\|_2 : y_{n+1} = -1\}$ . Solving this problem results in the nominal solution  $y^N$ . The condition number of  $\bar{A}$  is  $1.16 \times 10^5$ . This means that the nominal solution might be sensitive to an error in  $\bar{A}$ . Let us assume that the maximal inaccuracy in the coefficients of the first  $(n+1)$  columns of  $\bar{A}$  is 1%, and the aggregated inaccuracy in the temperature data cannot exceed 0.1% of the aggregated values. Hence, we consider the following uncertainty set:

$$\mathcal{Z} = \left\{ \Delta \in \mathbb{R}^{m \times (n+2)} : \left| \sum_{j \in \mathcal{J}} \sum_{i=1}^m \Delta_{ij} \right| \leq \rho, |\Delta_{ij}| \leq \bar{\rho}_j, \Delta_{i(n+2)} = 0, \begin{matrix} i = 1, \dots, m, \\ j = 1, \dots, n+1 \end{matrix} \right\},$$

where  $\mathcal{J}$  is the set containing the indices of the temperature columns and  $\rho = 0.001 \sum_{j \in \mathcal{J}} \sum_{i=1}^m |\bar{A}_{ij}|$ ,  $\bar{\rho}_j = 0.01 \max_{i=1, \dots, m} |\bar{A}_{ij}|$ ,  $j = 1, \dots, n+1$ . To obtain the value of  $\Omega$  in Assumption (A), we first notice that

$$\sup_{\Delta \in \mathcal{Z}} \|\Delta\|_{2,2} \leq \sup_{\substack{\Delta : \\ |\Delta_{ij}| \leq \bar{\rho}_j \\ i=1, \dots, m \\ j=1, \dots, n+1.}} \sup_{x: \|x\|_2=1} \|\Delta x\|_2 = \sup_{x: \|x\|_2=1} \sup_{\substack{\Delta : \\ |\Delta_{ij}| \leq \bar{\rho}_j \\ i=1, \dots, m \\ j=1, \dots, n+1.}} \|\Delta x\|_2 = \sup_{x: \|x\|_2=1} \sqrt{m} \|\bar{\rho}_j x_j\|_{j=1, \dots, n+1} \|1\|_1 \quad (43)$$



**Figure 5.** The empirical cumulative distribution functions (ECDF) of the absolute errors of two linear regressions: the solid blue line corresponds to  $y^I$  obtained from our inner approximation problem and the dashed red line corresponds to  $y^N$  obtained by solving (42). The left and right figures illustrate the ECDF of the absolute errors of the two linear regressions on the test dataset and the randomly generated scenarios, respectively.

and due to symmetry we can reformulate the far right optimization problem in (43) to

$$\sup_{x: \|x\|_2 \leq 1} \sqrt{m} \sum_{j=1}^{n+1} \bar{\rho}_j x_j.$$

So, we use

$$\Omega := \sqrt{m} \max_{x \in \mathbb{R}^{n+2}: \|x\|_2 \leq 1} \sum_{j=1}^{n+1} \bar{\rho}_j x_j.$$

For this instance, Assumption (B) does not hold. Therefore, we only consider the inner approximation

$$\min_{\substack{W \in \mathbb{R}^{(n+2) \times (n+2)} \\ y \in \mathbb{R}^{(n+2)} \\ \eta \in \mathbb{R}}} \left\{ \text{trace}((\bar{A}^T \bar{A} + \Omega^2 I_{n+2})W) + \delta_{\mathcal{Z}}^*(2\bar{A}W) + \frac{\eta}{4} : \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1}, \quad y_{n+1} = -1 \right\},$$

to obtain a robust solution  $y^I$ . We use 75% of the observations in the dataset to construct the regression model and the remaining ones to test the performance of the regression lines. To obtain the robust regression line, denoted by  $y^I$ , we solve the inner approximation, which takes around 276.5 seconds to be solved. Moreover, the nominal regression line, denoted by  $y^N$ , is obtained in around 1.1 seconds by solving (42). Figure 5 shows the empirical cumulative distribution function (ECDF) of the absolute errors of each regression lines on: (i) the rest of the 25% observations in the dataset, called the test dataset, and (ii) on 100 randomly generated scenarios  $\Delta$  from the uncertainty set  $\mathcal{Z}$ . We use each random scenario to construct a possible inaccuracy noise that is neglected in the data. As one can see in Figure 5, the performances of  $y^I$  and  $y^N$  are close while measured on the test dataset and randomly generated scenarios. The main distinction of the performances are in the skewness of the empirical distributions of the absolute errors. As can be seen, both distributions are right-skewed but the distribution of the absolute errors of the regression line  $y^N$  has a higher skewness.

Table 2 shows the comparison between the worst-case performances of  $y^I$  and  $y^N$ , where  $\Delta_I^*$  and  $\Delta_N^*$  are the worst-case scenarios for  $y^I$  and  $y^N$ , respectively, obtained using the heuristic

	$y^I$	$y^N$
$\ (A + \Delta_I^*)y\ _2$	13,714.6	15,104.0
$\ (A + \Delta_N^*)y\ _2$	13,559.7	15,272.3

**Table 2.** Performances of the inner approximation solution,  $y^I$ , and the nominal solution,  $y^N$ , on the worst-case scenarios.  $\Delta_I^*$  and  $\Delta_N^*$  are the scenarios generated in the uncertainty set using the algorithm in Appendix D for  $y^I$  and  $y^N$ , respectively.

algorithm proposed in Appendix D. As one can see,  $y^I$  has a better performance with respect to both worst-case scenarios  $\Delta_I^*$  and  $\Delta_N^*$ . More specifically, the regression line obtained from the inner approximation results in 10% improvement on the error realized in the worst-case scenarios. This shows the superiority of the robust regression line since the difference between its performances in the randomly generated inaccuracies as well as the test dataset are close to the ones delivered using the nominal solution.

## Acknowledgment

The research of the first author is partially supported by EU Marie Curie Initial Training Network number 316647 (“Mixed Integer Nonlinear Optimization (MINO)”), as well as by the 4TU strategic research and capacity building programme DeSIRE (Designing Systems for Informed Resilience Engineering), as part of the 4TU-programme High Tech for a Sustainable Future (HTSF). Furthermore, the authors would like to thank the anonymous reviewers and the associate editor for their helpful and constructive comments that greatly contributed to improving the paper.

## Appendices

The appendix of this paper consists of four parts. Part A contains the proofs of several lemmas and propositions. In the second part, we provide two simple examples to illustrate the results of Section 4. We provide different methods in Appendix C to check Assumptions (A) and (B). Finally, we propose a heuristic algorithm in Appendix D to find a worst-case scenario in the uncertainty set defined in (41) corresponding to a solution  $y$ .

## A Proofs

### A.1 Proof of Lemma 2

$$(i) \delta_{\mathcal{Z}}^*(U) = \sup_{\Delta \in \mathbb{R}^{n \times n}} \{ \text{trace}(\Delta U^T) : \text{vec}(\Delta) \in \mathcal{U} \} = \sup_{\Delta \in \mathbb{R}^{n \times n}} \{ \text{vec}(U)^T \text{vec}(\Delta) : \text{vec}(\Delta) \in \mathcal{U} \} = \delta_{\mathcal{U}}^*(\text{vec}(U)).$$

$$(ii) \delta_{\mathcal{Z}}^*(U) = \sup_{\Delta \in \mathcal{Z}} \{ \text{trace}(\Delta U^T) \} = \sup_{\zeta \in \mathcal{U}} \left\{ \sum_{i=1}^k \text{trace}(\zeta_i \Delta^i U^T) \right\} \\ = \sup_{\zeta \in \mathcal{U}} \left\{ \zeta^T [\text{trace}(\Delta^i U^T)]_{i=1, \dots, k} \right\} = \delta_{\mathcal{U}}^* \left( [\text{trace}(\Delta^i U^T)]_{i=1, \dots, k} \right).$$

$$(iii) \delta_{\mathcal{Z}}^*(U) = \sup_{\Delta \in \mathcal{Z}} \{ \text{trace}(\Delta U^T) \} = \sup_{\Theta \in \mathcal{U}} \{ \text{trace}(L \Theta R U^T) \} = \sup_{\Theta \in \mathcal{U}} \{ \text{trace}(\Theta R U^T L) \} = \delta_{\mathcal{U}}^*(L^T U R^T).$$

(iv) Let  $\bar{\Delta} \in \mathbb{R}^{n \times n}$  be such that  $L \circ \bar{\Delta} \in \mathcal{U}$ . If there exist  $\bar{i}, \bar{j} = 1, \dots, n$  for which  $L_{\bar{i}\bar{j}} = 0$  and  $U_{\bar{i}\bar{j}} \neq 0$ , then for any integer  $k$ , set

$$\Delta_{ij}^k := \begin{cases} \bar{\Delta}_{ij} & \text{if } i \neq \bar{i} \text{ or } j \neq \bar{j}, \\ k \text{sgn}(U_{\bar{i}\bar{j}}) \bar{\Delta}_{\bar{i}\bar{j}} & \text{otherwise,} \end{cases}$$

where  $\text{sgn}(\cdot)$  is the sign function. This constructs a sequence of matrices  $\{\Delta^k\}_{k \in \mathbb{N}}$  for which  $L \circ \Delta^k \in \mathcal{U}$  and  $\text{trace}(\Delta^k U^T)$  goes to  $+\infty$  when  $k$  tends to  $+\infty$ .

Now, assume  $U_{ij} = 0$  if  $L_{ij} = 0$ , for any  $i, j = 1, \dots, n$ . Then,

$$\begin{aligned}\delta_{\mathcal{Z}}^*(U) &= \sup_{L \circ \Delta \in \mathcal{U}} \sum_{i,j=1}^n \Delta_{ij} U_{ij} = \sup_{L \circ \Delta \in \mathcal{U}} \sum_{\substack{i,j=1,\dots,n: \\ L_{ij} \neq 0}} \Delta_{ij} U_{ij} = \sup_{L \circ \Delta \in \mathcal{U}} \sum_{\substack{i,j=1,\dots,n: \\ L_{ij} \neq 0}} L_{ij} \Delta_{ij} U_{ij} L_{ij}^\dagger \\ &= \sup_{L \circ \Delta \in \mathcal{U}} \text{trace} \left( L \circ \Delta (U \circ L^\dagger)^T \right) = \delta_{\mathcal{U}}^*(U \circ L^\dagger).\end{aligned}$$

$$(v) \quad \delta_{\mathcal{Z}}^*(U) = \sup_{\Delta \in \mathcal{Z}} \text{trace} (\Delta U^T) = \sup_{\substack{\Delta_i \in \mathcal{Z}^i \\ i=1,\dots,k}} \sum_{i=1}^k \text{trace} (\Delta^i U^T) = \sum_{i=1}^k \sup_{\Delta^i \in \mathcal{Z}^i} \text{trace} (\Delta^i U^T) = \sum_{i=1}^k \delta_{\mathcal{Z}^i}^*(U).$$

(vi) Similar to the proof of Lemma 9 in [4].

$$(vii) \quad \delta_{\mathcal{Z}}^*((U_1, \dots, U_k)) = \sup_{\Delta \in \mathcal{Z}} \text{trace} (\Delta (U_1, \dots, U_k)^T) = \sup_{\substack{\Delta_i \in \mathcal{Z}^i \\ i=1,\dots,k}} \text{trace} ((\Delta_1, \dots, \Delta_k)(U_1, \dots, U_k)^T) = \\ \sup_{\substack{\Delta_i \in \mathcal{Z}^i \\ i=1,\dots,k}} \text{trace} \left( \sum_{i=1}^k \Delta_i U_i^T \right) = \sum_{i=1}^k \sup_{\Delta_i \in \mathcal{Z}^i} \text{trace} (\Delta_i U_i^T) = \sum_{i=1}^k \delta_{\mathcal{Z}^i}^*(U_i).$$

$$(viii) \quad \delta_{\mathcal{Z}}^*(U) = \sup_{\Delta \in \mathcal{Z}} \text{trace} (\Delta U^T) = \sup_{\substack{\Delta^i \in \mathcal{Z}^i \\ \lambda_i \geq 0 \\ i=1,\dots,k}} \left\{ \sum_{i=1}^k \lambda_i \text{trace} (\Delta^i U^T) : \sum_{i=1}^k \lambda_i = 1 \right\} =$$

$$\max_{i=1,\dots,k} \sup_{\Delta^i \in \mathcal{Z}^i} \text{trace} (\Delta^i U^T) = \max_{i=1,\dots,k} \delta_{\mathcal{Z}^i}^*(U). \quad \square$$

## A.2 Proof of Lemma 3(b)

The assumptions imply that

$$\begin{aligned}\delta_{\mathcal{Z}}^*(U) &= \sup_{\Delta} \left\{ \text{trace} (\Delta U^T) : \Delta^l \preceq \Delta \preceq \Delta^u \right\} = \max_{\Delta} \left\{ \text{trace} \left( \frac{U + U^T}{2} \Delta \right) : \Delta^l \preceq \Delta \preceq \Delta^u \right\} \\ &= \min_{\Lambda_1, \Lambda_2} \left\{ \text{trace} (\Delta^u \Lambda_2) - \text{trace} (\Delta^l \Lambda_1) : \Lambda_2 - \Lambda_1 = \frac{U + U^T}{2}, \Lambda_1, \Lambda_2 \succeq 0_{n \times n} \right\},\end{aligned}$$

where the last equality holds because of conic duality (both problems are strictly feasible).  $\square$

## A.3 Proof of Lemma 4(ii)

Lemma 1(c) implies that  $\|U\|_{2,2}^2$  is the largest eigenvalue of  $UU^T$ . Hence,  $\|U\|_{2,2}^2 \leq \rho^2$  can be reformulated as  $UU^T \preceq \rho^2 I_n$ , which by using Schur Complement Lemma (see, e.g., Appendix A.5.5 in [16]) is equivalent to  $\begin{bmatrix} \rho^2 I_n & U \\ U^T & I_n \end{bmatrix} \succeq 0_{2n \times 2n}$ .  $\square$

## A.4 Proof of the statement in Example 1

$y \in \mathbb{R}^n$  satisfies (16) if and only if

$$y^T \bar{A} y + \sup_{\zeta \in \mathcal{Z}} \left\{ \zeta^T \left[ y^T A^i y + b^{i^T} y \right]_{i=1,\dots,t} \right\} + \bar{b}^T y + c \leq 0. \quad (44)$$

Now, we show that  $y \in \mathbb{R}^n$  satisfies (44) if and only if there exists  $v \in \mathbb{R}^t$  such that

$$y^T \bar{A} y + \sup_{\zeta \in \mathcal{Z}} \{ \zeta^T v \} + \bar{b}^T y + c \leq 0, \quad v \geq \left[ y^T A^i y + b^{i^T} y \right]_{i=1,\dots,t}. \quad (45)$$

It is clear that if  $y \in \mathbb{R}^n$  and  $v \in \mathbb{R}^t$  satisfy (45) then due to nonnegativity of  $\zeta \in \mathcal{Z}$ ,

$$\zeta^T v \geq \zeta^T \left[ y^T A^i y + b^{i^T} y \right]_{i=1,\dots,t},$$

which implies  $y \in \mathbb{R}^n$  satisfies (44). Now let  $y \in \mathbb{R}^n$  satisfies (44). Then setting  $v = \left[ y^T A^i y + b^{i^T} y \right]_{i=1,\dots,t}$  implies  $y \in \mathbb{R}^n$  and  $v \in \mathbb{R}^t$  satisfy (45), which can be reformulated as (17).  $\square$



## A.5 Proof of Proposition 3

Let  $(y, W)$  satisfy

$$\hat{\mu}^T y + \text{trace}(\hat{\Sigma}W) + \sqrt{\frac{\chi_{\text{rank}(V), 1-\alpha}^2}{n}} \left\| (\Psi R^{-1})^T \begin{pmatrix} y \\ \text{svec}(W) \end{pmatrix} \right\|_2 + c \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}.$$

We know that  $\sqrt{\chi_{d, 1-\alpha}^2} \geq z_{1-\alpha}$ , for any  $d$ . Also, we have

$$\text{trace}(W\hat{\Sigma}) \geq \text{trace}(yy^T\hat{\Sigma}) = y^T\hat{\Sigma}y,$$

where the inequality is because  $\hat{\Sigma} \succeq 0_{n \times n}$  and  $W \succeq yy^T$  using Schur Complement Lemma. Therefore,  $(y, W)$  satisfies

$$\hat{\mu}^T y + y^T\hat{\Sigma}y + \frac{z_{1-\alpha}}{\sqrt{n}} \left\| (\Psi R^{-1})^T \begin{pmatrix} y \\ \text{svec}(W) \end{pmatrix} \right\|_2 + c \leq 0, \quad \begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1},$$

which is the same as (34), since  $\Psi$  is diagonal.  $\square$

## B Some illustrative examples

**Example 3.** Let  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{n \times n} : \|\Delta\|_F \leq 1\}$  and let the assumptions of Theorem 1 hold. Then, using Theorem 1(I), Lemma 1, and Lemma 3(a),  $y$  satisfies (4a) if and only if there exists  $W \in \mathbb{R}^{n \times n}$  such that  $\text{trace}(\bar{A}W) + \bar{b}^T y + c + \|W + ya^T\|_F \leq 0$  and  $\begin{bmatrix} W & y \\ y^T & 1 \end{bmatrix} \succeq 0_{n+1 \times n+1}$ .  $\square$

In the next example, we derive a tractable reformulation of the RC in the form (4b) with the uncertainty set similar to the one proposed by [18].

**Example 4.** Consider the constraint

$$\sqrt{y^T A(\Delta)y + b(\zeta)^T y + c} \leq 0 \quad \forall (\zeta, \Delta) \in \mathcal{Z}, \quad (46)$$

where  $\zeta \in \mathbb{R}^n$ ,  $\Delta \in \mathbb{R}^{n \times n}$  are uncertain parameters,  $A(\Delta) = \bar{A} + \Delta$ ,  $b(\zeta) = \bar{b} + D\zeta$ ,  $\bar{A}, D \in \mathbb{R}^{n \times n}$ ,  $\bar{b} \in \mathbb{R}^n$ , and  $\mathcal{Z} = \mathcal{Z}_1 \cap \mathcal{Z}_2$ ,

$$\mathcal{Z}_1 = \left\{ (\zeta, \Delta) : \begin{bmatrix} 1 & \zeta^T \\ \zeta & \Delta \end{bmatrix} \succeq 0_{n+1 \times n+1} \right\}, \quad \mathcal{Z}_2 = \{(\zeta, \Delta) : \Delta^l \preceq \Delta \preceq \Delta^u\},$$

with given  $\Delta^l$  and  $\Delta^u$  such that  $\Delta^u - \Delta^l \succ 0_{n \times n}$ . Also, assume that the assumptions of Theorem 1 hold. By Lemma 2(vi),

$$\delta_{\mathcal{Z}}^*(U, v) = \min_{\substack{U^1, U^2 \in \mathbb{R}^{n \times n} \\ v^1, v^2 \in \mathbb{R}^n}} \{ \delta_{\mathcal{Z}_1}^*(U^1, v^1) + \delta_{\mathcal{Z}_2}^*(U^2, v^2) : U^1 + U^2 = U, v^1 + v^2 = v \}.$$

Following a similar line of reasoning as in the proof of Theorem 1(II),  $y \in \mathbb{R}^n$  satisfies (46) if and only if there exist  $W, U^1, U^2 \in \mathbb{R}^{n \times n}$ ,  $v^1, v^2 \in \mathbb{R}^n$  and  $\eta \in \mathbb{R}$  such that

$$\begin{cases} \text{trace}(\bar{A}W) + \bar{b}^T y + c + \delta_{\mathcal{Z}_1}^*(v^1, U^1) + \delta_{\mathcal{Z}_2}^*(v^2, U^2) + \frac{\eta}{4} \leq 0, \\ \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1}, U^1 + U^2 = W, v^1 + v^2 = D^T y. \end{cases} \quad (47)$$

Using Lemma 3(b), (47) is equivalent to

$$\begin{cases} \text{trace}(\bar{A}W) + \bar{b}^T y + c + \text{trace}(\Delta^u \Lambda_2) - \text{trace}(\Delta^l \Lambda_1) + \frac{\eta}{4} + \gamma \leq 0, \\ U^1 + U^2 = W, \quad \Lambda_2 - \Lambda_1 = \frac{U^2 + U^{2T}}{2}, \quad \Lambda_1, \Lambda_2 \succeq 0_{n \times n}, \\ \begin{bmatrix} W & y \\ y^T & \eta \end{bmatrix} \succeq 0_{n+1 \times n+1}, \quad \begin{bmatrix} \frac{U^1 + U^{1T}}{2} & \frac{1}{2} D^T y \\ \frac{1}{2} y^T D & -\gamma \end{bmatrix} \preceq 0_{n+1 \times n+1}, \end{cases}$$

for some  $\Lambda_1, \Lambda_2 \in S_n$  and  $\gamma \in \mathbb{R}$ .  $\square$

## C How to check Assumptions (A) and (B)

In this section, we provide methods that can be used to check Assumptions (A) and (B), each in a separate subsection.

### C.1 Finding $\Omega$ for which Assumption (A) holds

Assumption (A) states that there exists an upper bound  $\Omega$  for  $\sup_{\Delta \in \mathcal{Z}} \|\Delta\|_{2,2}$ . This assumption is equivalent to the boundedness of the uncertainty set  $\mathcal{Z}$ . So, checking this assumption can be done easily; however, in our inner approximation, we use the value of  $\Omega$ . Thus, in this section we provide methods to obtain it. Notice that  $\|\cdot\|_{2,2}$  is a convex function and the maximization of a convex function over a set, in general, is *NP-hard*. However, in Proposition 2, we show how to compute  $\Omega$  for the box uncertainty set.

In the cases for which  $\sup_{\Delta \in \mathcal{Z}} \|\Delta\|_{2,2}$  cannot be computed efficiently, one may use an upper bound for  $\|\cdot\|_{2,2}$  to calculate  $\Omega$ . For instance, one can approximate  $\mathcal{Z}$  with the union of simplices or boxes (see, e.g., [5] and [3]) and then find the maximum of  $\|\Delta\|_{2,2}$  over the union by means of Proposition 2 (for boxes) or by checking the vertices (for simplices).

### C.2 Checking Assumption (B)

Regarding Assumption (B), it is mentioned in Section 8.2 in [8] that finding a robust solution to an uncertain linear matrix inequality, in general, is *NP-hard*. Hence, there is no efficient way, in general, to check Assumption (B) exactly. However, there is much research that provides different methods to check Assumption (B). We refer the reader to the papers [19], [12], and Chapters 8 and 9 of the book [8]. Moreover, the problem may have specific characteristics from which this assumption can be certified. We refer the reader to the similar discussion in Section 4. In the following proposition, we provide an equivalent statement to Assumption (B) for uncertainty sets defined by matrix norms.

**Proposition 4.** *Let  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{n \times n} : \|\Delta\| \leq \rho\}$ , for a general matrix norm  $\|\cdot\|$ . Then Assumption (B) holds if and only if  $\sup_{y \in \mathbb{R}^n} \{-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y y^T\|^*\} = 0$ .*

*Proof.* Proof.

Assumption (B) holds if and only if

$$\begin{aligned}
 & \forall y \in \mathbb{R}^n \quad \forall \Delta \in \mathcal{Z} && y^T \bar{A}^T \bar{A} y + 2 \text{trace}(y y^T \bar{A}^T \Delta) \geq 0 \\
 \Leftrightarrow & \forall y \in \mathbb{R}^n && -y^T \bar{A}^T \bar{A} y + \sup_{\Delta \in \mathcal{Z}} \{-2 \text{trace}(y y^T \bar{A}^T \Delta)\} \leq 0 \\
 \Leftrightarrow & \forall y \in \mathbb{R}^n && -y^T \bar{A}^T \bar{A} y + \delta_{\mathcal{Z}}^*(-2 \bar{A} y y^T) \leq 0 \\
 \Leftrightarrow & \forall y \in \mathbb{R}^n && -y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y y^T\|^* \leq 0 \\
 \Leftrightarrow & && \sup_{y \in \mathbb{R}^n} \{-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y y^T\|^*\} \leq 0. \tag{48}
 \end{aligned}$$

Now, we show that, given  $\bar{A} \in \mathbb{R}^n$ , the optimal value of  $\sup_{y \in \mathbb{R}^n} \{-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y y^T\|^*\}$ , denoted by  $\tau^*$ , is either 0 or  $+\infty$ .

Clearly, the objective value of  $y = 0_n \in \mathbb{R}^n$  is 0. Hence,  $\tau^* \geq 0$ . If for any  $y \in \mathbb{R}^n$  the objective value is nonpositive, then  $\tau^* = 0$ . Now, let us assume that there exists a  $y \in \mathbb{R}^n$  such that the objective value is positive. Then for any  $\alpha \in \mathbb{R}$ , the objective value of  $\alpha y$  is

$$-(\alpha y)^T \bar{A}^T \bar{A} (\alpha y) + 2\rho \|\bar{A} (\alpha y) (\alpha y)^T\|^* = \alpha^2 (-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y y^T\|^*).$$

Hence, the objective value of  $\alpha y$  goes to  $+\infty$  when  $\alpha$  tends to  $+\infty$ . Therefore, the optimal value in this case is  $+\infty$ .  $\square$

Proposition 4 provides an unconstrained optimization problem equivalent to checking Assumption (B). In the next proposition, we show how one can use Proposition 4 to check Assumption (B) for box uncertainty sets.

**Proposition 5.** Let  $\mathcal{Z} = \{\Delta \in \mathbb{R}^{n \times n} : \|\Delta\|_\infty \leq \rho\}$ . For this uncertainty set Assumption (B) holds if and only if the following optimization problem has a nonnegative optimal value:

$$\min_{y \in \mathbb{R}^n} \{y^T \bar{A}^T \bar{A} y - 2\rho \|\bar{A} y\|_1 : \|y\|_1 = 1\}. \quad (49)$$

*Proof.* Proof. Using Proposition 4, Assumption (B) holds if and only if

$$\sup_{y \in \mathbb{R}^n} \{-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y\|_1\} = \sup_{y \in \mathbb{R}^n} \{-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y\|_1 \|y\|_1\} = 0. \quad (50)$$

As it is mentioned in the proof of Proposition 4, the optimal value of (50) is either zero or  $+\infty$ . Therefore, we know if Assumption (B) does not hold then the optimal value of (50) is  $+\infty$  and hence there exists  $\bar{y} \in \mathbb{R}^n$  such that

$$\bar{y}^T \bar{A}^T \bar{A} \bar{y} - 2\rho \|\bar{A} \bar{y}\|_1 \|\bar{y}\|_1 < 0.$$

Let us define  $\hat{y} := \frac{\bar{y}}{\|\bar{y}\|_1}$ . So, we have:

$$\hat{y}^T \bar{A}^T \bar{A} \hat{y} - 2\rho \|\bar{A} \hat{y}\|_1 = \hat{y}^T \bar{A}^T \bar{A} \hat{y} - 2\rho \|\bar{A} \hat{y}\|_1 \|\hat{y}\|_1 = \frac{1}{\|\bar{y}\|_1^2} (\bar{y}^T \bar{A}^T \bar{A} \bar{y} - 2\rho \|\bar{A} \bar{y}\|_1 \|\bar{y}\|_1) < 0.$$

So, in (49) the objective value of the feasible solution  $\hat{y}$  is negative. Hence, the optimal value of (49) is negative.

Now, let us assume that Assumption (B) holds. By Proposition 4 for any  $y \in \mathbb{R}^n$  we have that  $-y^T \bar{A}^T \bar{A} y + 2\rho \|\bar{A} y\|_1 \|y\|_1 \leq 0$ , for any  $y \in \mathbb{R}^n$ . Therefore, for any  $y$  in the set  $\mathcal{F} := \{y \in \mathbb{R}^n : \|y\|_1 = 1\}$ , we have  $y^T \bar{A}^T \bar{A} y - 2\rho \|\bar{A} y\|_1 \geq 0$  and hence (49) has a nonnegative optimal value.  $\square$

We emphasize that the optimization problem in (49) belongs to the class of DC (Difference of Convex) optimization problems, for which an extensive literature exists (see, e.g., [1, 31, 37]).

**Remark 9.** In this paper, we use (19) to check Assumption (B). If the optimal solution of (19) is nonnegative, then the inequality in (49) and hence Assumption (B) hold. Furthermore, (19) can be seen as an approximation of (49). The intuition behind this relaxation is that in (49) we want to minimize  $\|\bar{A} y\|_2$  and simultaneously maximizing  $\|\bar{A} y\|_1$ . Due to symmetrical behaviour of norm functions, we only restrict the problem to be optimized on  $\{y \in \mathbb{R}^n : \bar{A} y \geq 0, \|y\|_1 = 1\}$ , and then relax it into  $\{y \in \mathbb{R}^n : \bar{A} y \geq 0, \|y\|_1 \leq 1\}$ .  $\square$

**Remark 10.** For a general uncertainty set, using the  $\Omega$  from Assumption (A) and Proposition 1, we know Assumption (B) holds if  $\lambda_{\min}(\bar{A}^T \bar{A}) \geq 4\Omega^2$ , where  $\lambda_{\min}$  denotes the smallest eigenvalue of  $\bar{A}^T \bar{A}$ .  $\square$

## D A heuristic method to find a worst-case scenario in a norm approximation problem

In this section, we describe a heuristic method to find a worst-case scenario in the uncertainty set  $\mathcal{Z}$ , defined in (41), corresponding to a given solution  $y$ . We assume, without loss of generality, that the components of  $y$  are sorted such that their absolute values are descending.

The idea behind the algorithm comes from the fact that for the box uncertainty set

$$\{\Delta \in \mathbb{R}^{n \times n} : \|\Delta\|_\infty \leq \rho\}$$

and a solution  $y$ , the worst-case scenario  $\Delta^\square$  is defined by

$$\Delta_{ij}^\square = \text{sgn}([Ay - b]_i) \text{sgn}(y_j) \rho, \quad i, j = 1, \dots, n,$$

where  $\text{sgn}(\cdot)$  is the sign function. To see this, we first rewrite  $\|(\bar{A} + \Delta)y - b\|_2$  as  $\|\Delta y + \bar{A}y - b\|_2$  and we recall that the worst-case scenario is a vertex of the box uncertainty set, which are matrices

whose components are either  $\rho$  or  $-\rho$ . We know that  $\Delta$  maximizes  $\|\Delta y + \bar{A}y - b\|_2$  if for any  $i = 1, \dots, n$ , the value of  $(\Delta y)_i$  is the highest and its sign is the same as the sign of  $(\bar{A}y - b)_i$ . To make the value of  $(\Delta y)_i$  the highest for any  $i = 1, \dots, n$ , any component of  $\Delta_{ij}$  should have the same sign as  $y_j$ , with  $|\Delta_{ij}| = \rho$ ,  $j = 1, \dots, n$ . Hence,  $\Delta^\square$  is the maximizer of  $\|(A + \Delta)y - b\|_2$  over the box uncertainty set.

So, in the definition of  $\mathcal{Z}$ , if  $B^1 = B^2 = 0_{n \times n}$ , then  $\Delta^\square$  is the worst-case scenario. Also, if there exists  $i \in \{1, \dots, n\}$  such that the  $i$ th column of  $B^1$  and  $B^2$  are zero, then for any  $j = 1, \dots, n$ , the component in the  $j$ th row and  $i$ th column of the worst-case scenario is  $\Delta_{ij}^\square$ . So, from now on we assume that  $B^1$  and  $B^2$  do not have any zero columns in common.

Let  $\tilde{\Delta} = 0_{n \times n}$ . As  $y_1$  has the largest absolute value, changing the first column of  $\tilde{\Delta}$  may result in a high increase in the value of  $\|(\bar{A} + \tilde{\Delta})y - b\|_2$ . So, we start with  $j = 1$  and  $i = 1$ . If  $B_{ij}^1$  (or  $B_{ij}^2$ ) is 1, then we change  $\tilde{\Delta}_{ij}$  to  $\text{sgn}([\bar{A}y - b]_{i,n})\text{sgn}(y_j)\rho$ . We increase  $i$  by one and continue the same procedure. We also make sure that the number of changes happened because of  $B_{ij}^1 = 1$  (or  $B_{ij}^2 = 1$ ) does not exceed  $K$ . If we finish the procedure with  $i = n$ , then we increase  $j$  by one and reset  $i = 1$ . This procedure guarantees that the resulting scenario  $\tilde{\Delta}$  is feasible. In the next example, we illustrate how the algorithm works.

**Example 5.** Let  $n = 3$ ,

$$\bar{A} = \begin{bmatrix} 2.8 & 3.2 & 5.1 \\ -2.5 & 3.6 & 0 \\ -1.5 & 2.7 & 3.0 \end{bmatrix}, \quad B^1 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad B^2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \quad y = \begin{bmatrix} -3 \\ 2 \\ -1 \end{bmatrix},$$

$K = 5$ , and  $\rho = 0.2$ . For this example,  $\bar{A}y - b = [-9.1 \quad 11.7 \quad 5.9]^T$  and

$$\Delta^\square = \begin{bmatrix} 0.2 & -0.2 & 0.2 \\ -0.2 & 0.2 & -0.2 \\ -0.2 & 0.2 & -0.2 \end{bmatrix}.$$

Let  $C_1$  and  $C_2$  be the changes occurring because of  $B^1$  and  $B^2$ , respectively. For  $\tilde{\Delta}$ , we start with  $i, j = 1$ . Since both  $B_{11}^1 = B_{11}^2 = 1$ , so  $\tilde{\Delta}_{11} = 0.2$ ,  $C_1 = 1$ , and  $C_2 = 1$ . Since  $K = 5$ , we have  $C_1 \leq K$ , and  $C_2 \leq K$ . So, we continue and increase  $i$  to 2. We have  $B_{21}^2 = 1$  while  $B_{21}^1 = 0$ , so we change  $\tilde{\Delta}_{21} = -0.2$ , and  $C_2 = 2$ . By continuing this procedure, we end up with

$$\tilde{\Delta} = \begin{bmatrix} 0.2 & -0.2 & 0.2 \\ -0.2 & 0.2 & 0 \\ -0.2 & 0.2 & 0 \end{bmatrix}.$$

For this example, the worst-case value of  $\|(A + \Delta)y - b\|_2$  over  $\mathcal{Z}$  is 17.78 (obtained using SCIP 5.1 [23]) while  $\|(A + \tilde{\Delta})y - b\|_2 = 17.75$ .  $\square$

As one can also see from Example 5, the algorithm proposed in this section is a heuristic and the obtained scenario may not be the worst-case scenario.

## References

- [1] A. A. Ahmadi and G. Hall. DC decomposition of nonconvex polynomials with algebraic techniques. *Mathematical Programming*, 169(1):69–94, May 2018.
- [2] Mokhtar S Bazaraa, Hanif D Sherali, and Chitharanjan Marakada Shetty. *Nonlinear programming: theory and algorithms*. John Wiley & Sons, 2013.
- [3] A. Bemporad, C. Filippi, and F. D. Torrisi. Inner and outer approximations of polytopes using boxes. *Computational Geometry*, 27(2):151–178, 2004.
- [4] 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.

- [5] A. Ben-Tal, O. El Housni, and V. Goyal. A tractable approach for designing piecewise affine policies in two-stage adjustable robust optimization. *Mathematical Programming*, 2019.
- [6] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.
- [7] Aharon Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.
- [8] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.
- [9] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.
- [10] Aharon Ben-Tal and Arkadi Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [11] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *Mathematical Programming*, 167(2):235–292, Feb 2018.
- [12] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1):5–36, 2006.
- [13] Dimitris Bertsimas, Dessislava Pachamanova, and Melvyn Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.
- [14] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
- [15] J. M. Borwein and Q. J. Zhu. *Techniques of variational analysis*. CMS Books in Mathematics. Springer-Verlag New York, 2005.
- [16] Stephen Boyd and Lieven Vandenbergh. *Convex optimization*. Cambridge University Press, 2004.
- [17] L. M. Candanedo, V. Feldheim, and D. Deramaix. Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, 140:81–97, 2017.
- [18] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
- [19] L. El Ghaoui, F. Oustry, and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1):33–52, 1998.
- [20] Laurent El Ghaoui and Hervé Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.
- [21] Frank J Fabozzi, Petter N Kolm, Dessislava A Pachamanova, and Sergio M Focardi. *Robust portfolio optimization and management*. John Wiley & Sons, 2007.
- [22] M Fazel, H Hindi, and S P Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148)*, volume 6, pages 4734–4739. IEEE, 2001.
- [23] Ambros Gleixner, Leon Eifler, Tristan Gally, Gerald Gamrath, Patrick Gemander, Robert Lion Gottwald, Gregor Hendel, Christopher Hojny, Thorsten Koch, Matthias Miltenberger, Benjamin Mller, Marc E. Pfetsch, Christian Puchert, Daniel Rehfeldt, Franziska Schlsser, Felipe Serrano, Yuji Shinano, Jan Merlin Viernickel, Stefan Vigerske, Dieter Weninger, Jonas T. Witt, and Jakob Witzig. The SCIP optimization suite 5.0. *Optimization Online*, 2017.
- [24] Donald Goldfarb and Garud Iyengar. Robust convex quadratically constrained programs. *Mathematical Programming*, 97(3):495–515, 2003.
- [25] Bram L Gorissen and Dick den Hertog. Robust nonlinear optimization via the dual. 2015.
- [26] Bram L Gorissen, İhsan Yanıkoğlu, and Dick den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.

- [27] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.
- [28] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 2012.
- [29] V. Jeyakumar, G. Li, and J. Vicente-Pérez. Robust SOS-convex polynomial optimization problems: exact SDP relaxations. *Optimization Letters*, 9(1):1–18, 2015.
- [30] O. Linton. *Financial econometrics: models and methods*. Cambridge University Press, 2019.
- [31] Th. Lipp and S. Boyd. Variations and extension of the convex–concave procedure. *Optimization and Engineering*, 17(2):263–287, Jun 2016.
- [32] Johan Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, volume 3. Taipei, Taiwan, 2004.
- [33] MOSEK ApS. The MOSEK optimization toolbox for MATLAB manual, 2019.
- [34] P. M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization*, 1(1):15–22, Mar 1991.
- [35] Jason H Rife. The effect of uncertain covariance on a chi-square integrity monitor. *Navigation*, 60(4):291–303, 2013.
- [36] John Shawe-Taylor and Nello Cristianini. Estimating the moments of a random vector with applications. *Proceedings GRETSI 2003 Conference*, pages 1173–1178, 2003.
- [37] X. Shen, S. Diamond, Y. Gu, and S. Boyd. Disciplined convex-concave programming. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 1009–1014. IEEE, 2016.
- [38] L Trapani. Testing for (in)finite moments. *Journal of Econometrics*, 191(1):57–68, 2016.
- [39] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 2000.