

## LBS Research Online

D Bertsimas, D den Hertog, [J Pauphilet](#) and J Zhen  
Robust convex optimization: A new perspective that unifies and extends  
Article

This version is available in the LBS Research Online repository: <https://lbsresearch.london.edu/id/eprint/2617/>

Bertsimas, D, den Hertog, D, [Pauphilet, J](#) and Zhen, J  
(2023)

*Robust convex optimization: A new perspective that unifies and extends.*

Mathematical Programming, 200 (2). pp. 877-918. ISSN 0025-5610

DOI: <https://doi.org/10.1007/s10107-022-01881-w>

Springer Verlag (Germany)

<https://link-springer-com.lbs.idm.oclc.org/article...>

---

Users may download and/or print one copy of any article(s) in LBS Research Online for purposes of research and/or private study. Further distribution of the material, or use for any commercial gain, is not permitted.



# Robust convex optimization: A new perspective that unifies and extends

Dimitris Bertsimas<sup>1</sup> · Dick den Hertog<sup>2</sup> · Jean Pauphilet<sup>3</sup> · Jianzhe Zhen<sup>4</sup>

Received: 9 December 2021 / Accepted: 8 August 2022

© The Author(s) 2022

## Abstract

Robust convex constraints are difficult to handle, since finding the worst-case scenario is equivalent to maximizing a convex function. In this paper, we propose a new approach to deal with such constraints that unifies most approaches known in the literature and extends them in a significant way. The extension is either obtaining better solutions than the ones proposed in the literature, or obtaining solutions for classes of problems unaddressed by previous approaches. Our solution is based on an extension of the Reformulation-Linearization-Technique, and can be applied to general convex inequalities and general convex uncertainty sets. It generates a sequence of conservative approximations which can be used to obtain both upper- and lower- bounds for the optimal objective value. We illustrate the numerical benefit of our approach on a robust control and robust geometric optimization example.

**Keywords** Robust optimization · Convex analysis · Perspective functions · Linearization

**Mathematics Subject Classification** 90C25

---

✉ Dick den Hertog  
[d.denhertog@uva.nl](mailto:d.denhertog@uva.nl)

Dimitris Bertsimas  
[dbertsim@mit.edu](mailto:dbertsim@mit.edu)

Jean Pauphilet  
[jpauphilet@london.edu](mailto:jpauphilet@london.edu)

<sup>1</sup> Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, United States

<sup>2</sup> Amsterdam Business School, University of Amsterdam, Amsterdam, The Netherlands

<sup>3</sup> London Business School, London, United Kingdom

<sup>4</sup> swissQuant, Zürich, Switzerland

# 1 Introduction

In this paper, we consider a general hard robust constraint

$$h(A(x)z + b(x)) \leq 0, \quad \forall z \in \mathcal{Z}, \quad (1)$$

where  $h : \mathbb{R}^m \mapsto [-\infty, +\infty]$  is a proper, closed and convex function,  $A : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{m \times L}$ ,  $b : \mathbb{R}^{n_x} \mapsto \mathbb{R}^L$  are affine, and  $\mathcal{Z}$  is a nonempty convex subset of  $\mathbb{R}^L$ . Unlike inequalities that are concave in the uncertain parameters [4], a tractable *equivalent* reformulation for the robust constraint (1) is out of reach in general.

When  $h(\cdot)$  is (conic) quadratic, several exact reformulations have been proposed for specific uncertainty sets. We refer to [3, Chapter 6] for an overview of these methods. In the general case, generic approximate methods have been proposed for safely approximating (1). For homogeneous convex functions  $h(\cdot)$  and symmetric norm-based uncertainty sets  $\mathcal{Z}$ , Bertsimas and Sim [8] propose a computationally tractable safe approximation together with probabilistic guarantees. Zhen et al. [21] develop safe approximations for conic quadratic and SDP problems with polyhedral uncertainty sets. They first reformulate the robust inequality into an equivalent set of linear adaptive robust constraints, which can then be approximated by, e.g., static or affine decision rules. Roos et al. [18] extend their approach with affine decision rules to general convex functions  $h(\cdot)$  but still with polyhedral uncertainty sets. For general convex uncertainty sets, they propose a similar approach, yet using static decision rules only.

In this paper, we propose a new general safe approximation procedure for robust convex constraint (1). Our approach is ‘general’ in the sense that it can be applied to a generic convex function  $h(\cdot)$  and convex uncertainty set  $\mathcal{Z}$ . Using the Reformulation-Perspectification-Technique (RPT) as a unifying lens [20], our approach not only unifies many previous work from the literature but also extends them in a significant way. First, it is superior on special cases with homogeneous functions  $h(\cdot)$  and/or polyhedral uncertainty sets. Second, it is applicable to any convex function  $h(\cdot)$  and convex uncertainty set  $\mathcal{Z}$  under mild assumptions, namely that one can explicitly describe the uncertainty set and the domain of the conjugate of  $h$  (see definition in Sect. 1.2) via convex inequalities.

We develop our main approximation for a generic class of uncertainty sets, described as the intersection of a polyhedron and a convex set. We first prove that the original uncertain convex inequality (1) is equivalent to an uncertain linear constraint with a non-convex uncertainty set. In other words, we reformulate (1) as a linear robust constraint where the new uncertainty set  $\Theta$  is non-convex due to the presence of bilinear equality constraints. We then use the recently developed Reformulation-Perspectification-Technique (RPT) [20] to obtain a hierarchy of safe approximations, i.e, we construct convex sets  $\Theta_i$ ,  $i = 0, 1, \dots, 3$ , such that  $\Theta \subseteq \Theta_3 \subseteq \Theta_2 \subseteq \Theta_1 \subseteq \Theta_0$ . We show that this approach unifies a variety of methods known in the literature, and extends them in a significant way. When the uncertainty set lies in the non-negative orthant, we show that the first-level approximation  $\Theta_1$  coincides with the perspectification approach, a generalization of the approach of Bertsimas and Sim [8] for which we also provide probabilistic guarantees. Consequently, our approach connects and

generalizes further the work of Bertsimas and Sim [8] and Roos et al. [18] to a broader class of uncertainty sets, and improves them by proposing tighter approximations,  $\Theta_2$  and  $\Theta_3$ . For polyhedral uncertainty sets, our RPT-based approach with  $\Theta_1$  coincides with the approximation from Roos et al. [18] with linear decision rules [18, Theorem 2].

The paper is structured as follows: In Sect. 2, we present our RPT-based hierarchy of safe approximations. We connect it with existing results from the literature in Sect. 3. In particular, for uncertainty sets lying in the non-negative orthant, we show in Appendix A that our first-level approximation with  $\Theta_1$  can be alternatively derived using a ‘perspectification’ technique and provide probabilistic guarantees in this case. When applied to a minimization problem, our safe approximations provide a valid upper bound on the objective value. Our approach can also be adapted to derive lower bounds, as described in Sect. 4. Finally, we assess the numerical performance of our technique in Sect. 5, and conclude our findings in Sect. 6.

## 1.1 Examples

In this section, we present a few examples of the robust convex constraint (1) often encountered in the literature and summarize them in Table 1.

**Quadratic Optimization (QO)** We consider the general quadratic constraint

$$\mathbf{z}^\top \mathbf{F}(\mathbf{x})^\top \mathbf{F}(\mathbf{x}) \mathbf{z} + \mathbf{f}(\mathbf{x})^\top \mathbf{z} \leq g(\mathbf{x}), \quad \forall \mathbf{z} \in \mathcal{Z},$$

where  $\mathbf{F} : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{m \times L}$ ,  $\mathbf{f} : \mathbb{R}^{n_x} \mapsto \mathbb{R}^L$  and  $g : \mathbb{R}^{n_x} \mapsto \mathbb{R}$  are affine in  $\mathbf{x}$ . The constraint is of the form (1) with

$$h\left(\begin{bmatrix} \mathbf{y} \\ \bar{\mathbf{y}} \end{bmatrix}\right) := \mathbf{y}^\top \mathbf{y} + \bar{\mathbf{y}}, \quad \mathbf{A}(\mathbf{x}) := \begin{bmatrix} \mathbf{F}(\mathbf{x}) \\ \mathbf{f}(\mathbf{x})^\top \end{bmatrix}, \quad \text{and } \mathbf{b}(\mathbf{x}) := \begin{bmatrix} \mathbf{0} \\ g(\mathbf{x}) \end{bmatrix}.$$

Alternatively, the constraint can be represented as a second-order cone constraint

$$\left\| \begin{bmatrix} \mathbf{F}(\mathbf{x}) \mathbf{z} \\ (1 + \mathbf{f}(\mathbf{x})^\top \mathbf{z} - g(\mathbf{x}))/2 \end{bmatrix} \right\|_2 \leq (1 - \mathbf{f}(\mathbf{x})^\top \mathbf{z} + g(\mathbf{x})), \quad \forall \mathbf{z} \in \mathcal{Z},$$

which is also of the form (1) for

$$h\left(\begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}} \\ \bar{\mathbf{y}} \end{bmatrix}\right) := \sqrt{\mathbf{y}^\top \mathbf{y} + \tilde{\mathbf{y}}^2} + \bar{\mathbf{y}}, \quad \mathbf{A}(\mathbf{x}) := \begin{bmatrix} \mathbf{F}(\mathbf{x}) \\ \frac{1}{2} \mathbf{f}(\mathbf{x})^\top \\ \frac{1}{2} \mathbf{f}(\mathbf{x})^\top \end{bmatrix}, \quad \text{and } \mathbf{b}(\mathbf{x}) := \begin{bmatrix} \mathbf{0} \\ \frac{1-g(\mathbf{x})}{2} \\ -\frac{1+g(\mathbf{x})}{2} \end{bmatrix}.$$

**Piecewise linear constraints (PWL)** We consider the piece-wise linear convex constraint

$$\max_{k \in [m]} \left\{ \mathbf{a}_k(\mathbf{x})^\top \mathbf{z} + b_k(\mathbf{x}) \right\} \leq 0, \quad \forall \mathbf{z} \in \mathcal{Z},$$

**Table 1** Example of functions  $h(\cdot)$  and their corresponding conjugate, defined as  $h^*(\mathbf{w}) := \sup_{\mathbf{y} \in \text{dom } h} \{\mathbf{y}^\top \mathbf{w} - h(\mathbf{y})\}$ 

Type	$h$	$\text{dom } h^*$	$h^*$
QO 1	$h(\mathbf{y}, \bar{y}) := \mathbf{y}^\top \mathbf{y} + \bar{y}$	$\{(\mathbf{w}, \bar{w}) : \bar{w} = 1\}$	$\frac{1}{4} \mathbf{w}^\top \mathbf{w}$
QO 2	$h(\mathbf{y}, \bar{y}) := \ \mathbf{y}\ _2 + \bar{y}$	$\{(\mathbf{w}, \bar{w}) : \ \mathbf{w}\ _2 \leq 1, \bar{w} = 1\}$	0
PWL	$h(\mathbf{y}) := \max_i y_i$	$\{\mathbf{w} \geq 0 : \sum_i w_i = 1\}$	0
SML	$h(\mathbf{y}) := \sum_k \max_{i \in I_k} y_i$	$\left\{(\mathbf{w}_k)_k : \forall k, \begin{matrix} \mathbf{w}_k \geq 0, \\ \sum_{i \in I_k} w_{ki} = 1 \end{matrix}\right\}$	0
GO	$h(\mathbf{y}) = \log(\sum_i e^{y_i})$	$\{\mathbf{w} \geq 0 : \sum_i w_i = 1\}$	$\sum_i w_i \log w_i$
SC	$h(\mathbf{y}) := \sum_{i \in [I]} h_i(\mathbf{y})$	$\left\{(\mathbf{w}_i)_i : \begin{matrix} \sum_i \mathbf{w}_i = \mathbf{w}, \\ \mathbf{w}_i \in \text{dom } h_i^*, \forall i \end{matrix}\right\}$	$\min_{\mathbf{w}_i} \sum_i h_i^*(\mathbf{w}_i)$

where  $\mathbf{a}_k : \mathbb{R}^{n_x} \mapsto \mathbb{R}^L$  and  $b_k : \mathbb{R}^{n_x} \mapsto \mathbb{R}$  are affine functions of  $\mathbf{x}$ . Such a constraint is a special case of (1) with

$$h(\mathbf{y}) := \max_{k \in [m]} y_k, \quad \mathbf{A}(\mathbf{x}) := [\mathbf{a}_k(\mathbf{x})^\top]_k, \quad \text{and} \quad \mathbf{b}(\mathbf{x}) := [b_k(\mathbf{x})]_k.$$

**Sum-of-max of linear constraints (SML)** The sum-of-max of linear constraints is defined as  $h(\mathbf{y}) := \sum_k \max_{i \in I_k} y_i$ , for any  $\mathbf{y} \in \mathbb{R}^{n_y}$ .

**Geometric optimization (GO)** In geometric optimization, the function  $h$  in (1) is the log-sum-exp function defined by  $h(\mathbf{y}) := \log(e^{y_1} + \dots + e^{y_{n_y}})$  for any  $\mathbf{y} \in \mathbb{R}^{n_y}$ .

**Sum-of-convex constraints (SC)** The sum of general convex functions is defined as  $h(\mathbf{y}) := \sum_{i \in [I]} h_i(\mathbf{y})$ , where  $h_i : \mathbb{R}^m \mapsto [-\infty, +\infty]$ ,  $i \in [I]$ , are proper, closed and convex with  $\bigcap_{i \in [I]} \text{ri}(\text{dom } h_i) \neq \emptyset$ .

In this paper, we study robust convex constraints, and not only robust conic constraints, on purpose. It is generally believed that most convex problems can be modeled in terms of the five basic cones: linear, quadratic, semi-definite, power, and exponential cones. Still, we do not restrict our attention to robust conic optimization for the following reason: Although conic optimization is a powerful modeling framework, the uncertainty is often split over multiple constraints, and the robust version of the conic representation is not equivalent to the original one. For example, while the sum-of-max of linear functions is linear-cone representable, its robust counterpart is not necessarily equivalent with the robust version of the original constraint. The conclusion is that the right order is to first develop the robust counterpart of the nonlinear inequality and then reformulate it into a conic representation, instead of the other way around.

## 1.2 Notations

We use nonbold face characters ( $x$ ) to denote scalars, lowercase bold faced characters ( $\mathbf{x}$ ) to denote vectors, uppercase bold faced characters ( $\mathbf{X}$ ) to denote matrices, and bold calligraphic characters such as  $\mathcal{X}$  to denote sets. We denote by  $\mathbf{e}_i$  the unit vector with 1 at the  $i$ th coordinate and zero elsewhere, with dimension implied by the context.

We use  $[n]$  to denote the finite index set  $[n] = \{1, \dots, n\}$  with cardinality  $|[n]| = n$ . Whenever the domain of an optimization variable is omitted, it is understood to be the entire space (whose definition will be clear from the context).

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

A function  $f : \mathbb{R}^{n_y} \mapsto [-\infty, +\infty]$  is said to be *proper* if there exists at least one vector  $\mathbf{y}$  such that  $f(\mathbf{y}) < +\infty$  and for any  $\mathbf{y} \in \mathbb{R}^{n_y}$ ,  $f(\mathbf{y}) > -\infty$ . For a proper, closed and convex function  $f$ , we define its *conjugate* as  $f^*(\mathbf{w}) = \sup_{\mathbf{y} \in \text{dom } f} \{\mathbf{w}^\top \mathbf{y} - f(\mathbf{y})\}$ . When  $f$  is closed and convex,  $f^{**} = f$ . If the function  $f : \mathbb{R}^{n_y} \mapsto [-\infty, +\infty]$  is proper, closed and convex, the *perspective function* of  $f$  is defined for all  $\mathbf{y} \in \mathbb{R}^{n_y}$  and  $t \in \mathbb{R}_+$  as

$$\text{persp}_f : (\mathbf{y}, t) \mapsto \begin{cases} tf(\mathbf{y}/t), & \text{if } t > 0, \\ f_\infty(\mathbf{y}), & \text{if } t = 0. \end{cases}$$

The function  $f_\infty : \mathbb{R}^{n_y} \times \mathbb{R}_+ \mapsto [-\infty, +\infty]$  is the *recession function*, which can be equivalently defined as

$$f_\infty(\mathbf{y}) = \lim_{t \rightarrow 0} tf(\mathbf{y}/t) = \delta^*(\mathbf{y} | \text{dom } f^*).$$

Among others, the perspective and conjugate functions of  $f$  satisfy the relationship

$$tf\left(\frac{\mathbf{y}}{t}\right) = \sup_{\mathbf{w} \in \text{dom } f^*} \left\{ \mathbf{y}^\top \mathbf{w} - tf^*(\mathbf{w}) \right\} \quad \text{for } t > 0.$$

For ease of exposition, we use  $tf(\mathbf{y}/t)$  to denote the perspective function  $f$ .

## 2 The reformulation-perspectification approach

In this section, we describe our main approach, which is based on an extension of the Reformulation-Linearization-Technique, that is, Reformulation-Perspectification-Technique [20].

Our approach comprises three steps:

- **Step 1.** Reformulate the robust convex constraint (1) into a robust linear optimization problem with *bilinear* equalities in the uncertainty set.
- **Step 2.** Apply the Reformulation-Perspectification-Technique to get a safe approximation of the robust convex inequality that is linear in the uncertain parameters.
- **Step 3.** Construct a computationally tractable robust counterpart of the approximation obtained in Step 2 by using the approach described in Ben-Tal et al. [4].

## 2.1 Step 1: Bilinear reformulation

We first describe **Step 1** in more detail. The following proposition shows that the robust convex constraint (1) is equivalent to a robust linear constraint with bilinear equality constraints in the uncertainty set.

**Proposition 1** (Bilinear Reformulation) *The robust convex constraint (1) is equivalent to*

$$\sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \right\} \leq 0, \quad (2)$$

where

$$\Theta = \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \mid \mathbf{z} \in \mathcal{Z}, \mathbf{w} \in \text{dom } h^*, h^*(\mathbf{w}) \leq w_0, \mathbf{V} = \mathbf{w}\mathbf{z}^\top, \mathbf{v}_0 = w_0\mathbf{z} \right\}.$$

**Proof** Because  $h$  is closed and convex we have

$$h(s) = h^{**}(s) = \sup_{\mathbf{w} \in \text{dom } h^*} \left\{ s^\top \mathbf{w} - h^*(\mathbf{w}) \right\},$$

and thus

$$\begin{aligned} \sup_{\mathbf{z} \in \mathcal{Z}} h(\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x})) &= \sup_{\mathbf{z} \in \mathcal{Z}} \sup_{\mathbf{w} \in \text{dom } h^*} \mathbf{z}^\top \mathbf{A}(\mathbf{x})^\top \mathbf{w} + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - h^*(\mathbf{w}) \\ &= \sup_{\substack{\mathbf{z} \in \mathcal{Z} \\ \mathbf{w} \in \text{dom } h^* \\ h^*(\mathbf{w}) \leq w_0}} \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{w}\mathbf{z}^\top) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \\ &= \sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta} \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0, \end{aligned} \quad (3)$$

for the properly defined set  $\Theta$ .  $\square$

The complication in (2) is that the extended uncertainty set  $\Theta$  is not convex due to the bilinear constraint  $\mathbf{V} = \mathbf{w}\mathbf{z}^\top$ . Therefore, in **Step 2** we propose a safe approximation of  $\Theta$  based on the Reformulation-Perspectification-Technique [20]. Formally, we provide tractable approximations of the robust convex constraint in (1) by exhibiting *convex* sets  $\Phi$  such that  $\Theta \subseteq \Phi$ . Indeed, if  $\Theta \subseteq \Phi$ , then

$$\sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Phi} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \right\} \leq 0$$

constitutes a safe approximation of (1). In addition, this safe approximation is linear in the uncertain parameter  $(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0)$  and convex in the decision variable  $\mathbf{x}$ . Hence, provided that the new uncertainty sets  $\Phi$  is convex, its robust counterpart can be derived in a tractable manner using techniques from Ben-Tal et al. [4] (**Step 3**). Observe that since the objective function is linear, we can replace  $\Theta$  by  $\text{conv}(\Theta)$  in (2) without loss of optimality. So we effectively need to construct sets  $\Phi$  that approximate  $\text{conv}(\Theta)$  as closely as possible.

## 2.2 Step 2: Hierarchy of safe approximations

To develop a hierarchy of safe approximations (**Step 2**) for the robust non-convex constraint (2), we assume that the uncertainty set  $\mathcal{Z}$  is described through  $K_0$  linear and  $K_1$  convex inequalities.

**Assumption 1** The uncertainty set  $\mathcal{Z} = \{z \in \mathbb{R}^L \mid D z \leq d, c_k(z) \leq 0, k \in [K_1]\}$ , where  $D \in \mathbb{R}^{K_0 \times L}$ ,  $d \in \mathbb{R}^{K_0}$ , and  $c_k : \mathbb{R}^L \mapsto [-\infty, +\infty]$  is proper, closed and, convex for each  $k \in [K_1]$ .

Note that the decomposition of  $\mathcal{Z}$  into linear and convex constraints in Assumption 1 is not unique. Indeed, affine functions are also convex so one could assume  $K_0 = 0$  without loss of generality. However, the presence of linear constraints in the definition of  $\mathcal{Z}$  will be instrumental in deriving good approximations. Consequently, from a practical standpoint, we encourage to include as many linear constraints as possible in the linear system of inequalities  $D z \leq d$ . We illustrate this modeling choice on a simple example.

**Example 1** Consider the polytope  $\mathcal{Z} = \{z \in \mathbb{R}_+^L \mid \sum_{\ell=1}^L z_\ell \leq 1, \ell \in [L]\}$ .  $\mathcal{Z}$  satisfies Assumption 3 with  $K = 1$  and  $c_1(z) = \sum_{\ell} z_\ell - 1$ . Therefore,  $\mathcal{Z}$  also satisfies Assumption 1 with  $D = -I_L$ ,  $d = \mathbf{0}_L$ , and  $c_1(z) = \sum_{\ell} z_\ell - 1$ , hence  $K_0 = L$  and  $K_1 = 1$ . Alternatively,  $\mathcal{Z}$  satisfies Assumption 1 with  $K_0 = L + 1$  and  $K_1 = 0$  with

$$D = \begin{bmatrix} -I_L \\ e^\top \end{bmatrix}, \text{ and } d = \begin{bmatrix} \mathbf{0}_L \\ 1 \end{bmatrix}.$$

Similarly, we assume that  $\text{dom } h^*$  can be described with  $J_0$  linear and  $J_1$  convex inequalities.

**Assumption 2** The set  $\text{dom } h^* = \{w \mid F w \leq f, g_j(w) \leq 0, j \in [J_1]\}$  has  $J_0$  linear inequalities defined by  $F$  and  $f$ , and  $J_1$  nonlinear inequalities defined by  $g_j, j \in [J_1]$ , where  $g_j$  is proper, closed, and convex for each  $j \in [J_1]$ .

Under Assumptions 1 and 2, the non-convex uncertainty set  $\Theta$  can be represented as

$$\Theta = \left\{ (w, w_0, z, V, v_0) \mid \begin{array}{l} D z \leq d, c_k(z) \leq 0, k \in [K_1] \quad (z \in \mathcal{Z}) \\ F w \leq f, g_j(w) \leq 0, i \in [J_1] \quad (w \in \text{dom } h^*) \\ h^*(w) \leq w_0 \quad (\text{epigraph}) \\ V = w z^\top \\ v_0 = w_0 z \end{array} \right\}.$$

**Proposition 2** (Convex Relaxation via RPT) *For any  $(w, w_0, z, V, v_0) \in \Theta$ , we have*

- $\frac{d_k w - V D^\top e_k}{d_k - e_k^\top D z} \in \text{dom } h^*, \text{ for all } k \in [K_0],$
- $\frac{f_j z - F V e_j}{f_j - e_j^\top F w} \in \mathcal{Z}, \text{ for all } j \in [J_0].$



Each of these two set memberships can be expressed through constraints that are convex in  $(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0)$ . As a result, the three sets

$$\begin{aligned}\Theta_0 &= \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \left| \begin{array}{l} \mathbf{z} \in \mathcal{Z}, \\ \mathbf{w} \in \text{dom } h^*, \\ h^*(\mathbf{w}) \leq w_0 \end{array} \right. \right\}, \\ \Theta_1 &= \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \left| \begin{array}{l} (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta_0, \\ \frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \in \text{dom } h^*, \quad k \in [K_0] \end{array} \right. \right\}, \\ \Theta_2 &= \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \left| \begin{array}{l} (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta_1, \\ \frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \in \mathcal{Z}, \quad j \in [J_0] \end{array} \right. \right\}\end{aligned}$$

are convex and satisfy  $\Theta \subseteq \Theta_2 \subseteq \Theta_1 \subseteq \Theta_0$ .

**Proof** a) Consider a linear constraint on  $\mathbf{z}$ ,  $\mathbf{e}_k^\top (\mathbf{d} - \mathbf{D} \mathbf{z}) \geq 0$  for some  $k \in [K_0]$ . If  $d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z} > 0$ , algebraic manipulations yield

$$\mathbf{w} = \frac{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \mathbf{w} = \frac{d_k \mathbf{w} - (\mathbf{e}_k^\top \mathbf{D} \mathbf{z}) \mathbf{w}}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} = \frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}},$$

where the last equality follows from the fact that  $\mathbf{V} = \mathbf{w} \mathbf{z}^\top$ . Since  $\mathbf{w} \in \text{dom } h^*$ , we must have  $\frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \in \text{dom } h^*$ . We then derive convex constraints in  $(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0)$  that enforces this set membership. By multiplying both sides of a generic convex constraint on  $\mathbf{w}$ ,  $g(\mathbf{w}) \leq 0$ , by the non-negative scalar  $d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}$ , we obtain

$$(d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) g(\mathbf{w}) = (d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) g\left(\frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}}\right) \leq 0, \quad (4)$$

which is convex in  $(d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}, d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k)$ , since the constraint function is the perspective of the convex function  $g$ . Hence, it is convex in  $(\mathbf{w}, \mathbf{z}, \mathbf{V})$ . Note that if  $d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}$  is not strictly positive, constraint (4) should still hold by a limit argument. We apply this methodology to:

- Linear constraints on  $\mathbf{w}$ ,  $g(\mathbf{w}) = \mathbf{F} \mathbf{w} - \mathbf{f} \leq \mathbf{0}$ . In this case, (4) leads to

$$(d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) \left[ \mathbf{F} \frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} - \mathbf{f} \right] \leq \mathbf{0},$$

and can be expressed as a linear constraint in  $(\mathbf{w}, \mathbf{z}, \mathbf{V})$ :

$$\mathbf{F} \left[ d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k \right] \leq (d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) \mathbf{f}. \quad (5)$$

–  $g(\mathbf{w}) = g_j(\mathbf{w})$ ,  $j \in [J_1]$ , i.e.,

$$(d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) g_j \left( \frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \right) \leq 0, \quad \forall j \in [J_1], \quad (6)$$

– and the epigraph constraint,  $h^*(\mathbf{w}) \leq w_0$ , yielding

$$\begin{aligned} & (d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) \left[ h^* \left( \frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \right) - w_0 \right] \leq 0, \\ \iff & (d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}) h^* \left( \frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \right) \leq d_k w_0 - \mathbf{e}_k^\top \mathbf{D} \mathbf{v}_0, \end{aligned} \quad (7)$$

where the equivalence follows from  $\mathbf{v}_0 = w_0 \mathbf{z}$ .

Together, these constraints (5)–(6)–(7) enforce that

$$\frac{d_k \mathbf{w} - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{d_k - \mathbf{e}_k^\top \mathbf{D} \mathbf{z}} \in \text{dom } h^*.$$

b) Similarly, we can consider a linear constraint in  $\mathbf{w}$ ,  $f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w} \geq 0$ , for some  $j \in [J_0]$ . We have

$$\mathbf{z} = \frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \text{ so that } \frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \in \mathcal{Z}.$$

Given a generic convex constraint in  $\mathbf{z}$ ,  $c(\mathbf{z}) \leq 0$ , we obtain a new convex constraint

$$(f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}) c \left( \frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \right) \leq 0.$$

Applying these manipulations to the linear constraints  $\mathbf{D} \mathbf{z} \leq \mathbf{d}$  yields

$$\mathbf{D} [f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j] \leq (f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}) \mathbf{d}. \quad (8)$$

Taking  $c(\mathbf{z}) = c_k(\mathbf{z})$ ,  $k \in [K_1]$  we get

$$(f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}) c_k \left( \frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \right) \leq 0, \quad \forall k \in [K_1]. \quad (9)$$

Eventually, we obtain constraints (8)–(9), enforcing that

$$\frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \in \mathcal{Z}.$$

□

Proposition 2 exhibits a hierarchy of approximations to convexify the non-convex set  $\Theta$ . A safe approximation is then obtained by considering the robust linear constraint

$$\sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0) \in \Phi} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \right\} \leq 0, \quad (10)$$

with  $\Phi = \Theta_i$ ,  $i = 0, 1, 2$ . Observe that for  $\Theta_0$ , the robust counterpart of the safe approximation is

$$\mathbf{A}(\mathbf{x}) = \mathbf{0}, \quad h(\mathbf{b}(\mathbf{x})) \leq 0.$$

Accordingly, it is fair to admit that  $\Theta_0$  offers a poor approximation and that relevant safe approximations are to be found at higher orders of the hierarchy. A higher order in the hierarchy, however, does not necessarily imply a tighter safe approximation, as discussed in Sect. 2.5.

So far, we obtained convex constraints on  $(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0)$  by multiplying a linear constraint on  $\mathbf{z}$  (resp.  $\mathbf{w}$ ) with a convex constraint on  $\mathbf{w}$  (resp.  $\mathbf{z}$ ). However, in many cases multiplying two general convex constraints, e.g., multiplying  $c_k(\mathbf{z}) \leq 0$  with  $g_j(\mathbf{w}) \leq 0$ , also yields valuable convex constraints. The resulting set, which we denote  $\Theta_3$  and concisely define as

$$\Theta_3 = \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0) \mid \begin{array}{l} (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0) \in \Theta_2, \\ c_k(\mathbf{z}) \times g_j(\mathbf{w}) \leq 0, \quad k \in [K_1], \quad j \in [J_1] \end{array} \right\}$$

provides an even tighter safe approximation of (2). In the definition above, “ $c_k(\mathbf{z}) \times g_j(\mathbf{w}) \leq 0$ ” denotes valid convex inequalities obtained when multiplying “ $c_k(\mathbf{z}) \leq 0$ ” with “ $g_j(\mathbf{w}) \leq 0$ ”. In the following example, we illustrate this approach when  $c_k$  and  $g_j$  are two conic quadratic functions, which typically occurs with ellipsoidal uncertainty sets and a conic quadratic function  $h$ . We refer to Zhen et al. [20] for a thorough analysis of all 15 possible multiplications of any two of the five basic convex cones (linear, conic quadratic, power, exponential, semi-definite).

**Example 2** Consider two conic quadratic inequalities

$$\begin{cases} b_1 - \mathbf{a}_1^\top \mathbf{w} \geq \|\mathbf{D}_1 \mathbf{w} + \mathbf{p}_1\| \\ b_2 - \mathbf{a}_2^\top \mathbf{z} \geq \|\mathbf{D}_2 \mathbf{z} + \mathbf{p}_2\|, \end{cases}$$

the first one coming from the description of  $\text{dom } h^*$ , and the second is one of the constraints that define  $\mathcal{Z}$ . We now apply RPT to these inequalities, i.e., we multiply the LHS and RHS of the two constraints, and we multiply the LHS of the first constraint

with the second constraint, and multiply the LHS of the second constraint with the first constraint. We obtain:

$$\begin{aligned} & \begin{cases} (b_1 - a_1^\top w)(b_2 - a_2^\top z) \geq \|D_1 w + p_1\| \|D_2 z + p_2\| \\ (b_1 - a_1^\top w)(b_2 - a_2^\top z) \geq (b_1 - a_1^\top w) \|D_2 z + p_2\| \\ (b_1 - a_1^\top w)(b_2 - a_2^\top z) \geq (b_2 - a_2^\top z) \|D_1 w + p_1\| \end{cases} \\ \Leftrightarrow & \begin{cases} b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top w z^\top a_2 \geq \|(D_1 w + p_1)(D_2 z + p_2)^\top\|_2 \\ b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top w z^\top a_2 \geq \|b_1 D_2 z + b_1 p_2 - D_2 w z^\top a_1 - a_1^\top w p_2\| \\ b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top w z^\top a_2 \geq \|b_2 D_1 w + b_2 p_1 - D_1 w z^\top a_2 - a_2^\top z p_1\| \end{cases} \\ \Leftrightarrow & \begin{cases} b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top V a_2 \geq \|D_1 V D_2^\top + p_1 z^\top D_2^\top + D_1 w p_2^\top + p_1 p_2^\top\|_2 \\ b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top V a_2 \geq \|b_1 D_2 z + b_1 p_2 - D_2 V a_1 - a_1^\top w p_2\| \\ b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top V a_2 \geq \|b_2 D_1 w + b_2 p_1 - D_1 V a_2 - a_2^\top z p_1\|, \end{cases} \end{aligned}$$

since  $V = w z^\top$ . Here,  $\|\cdot\|_2$  denotes the 2-norm of a matrix defined as its largest singular value, i.e.,  $\|M\|_2 := \sigma_{\max}(M) = \sqrt{\lambda_{\max}(M M^\top)}$ . In particular, we used the fact that for rank-1 matrices  $M = u v^\top$ ,  $\|M\|_2 = \|u\| \|v\|$ . In practice, constraints of the form  $\|M\|_2 \leq \lambda$  can be modeled as semidefinite constraints using Schur complement. To improve scalability, however, one can replace the 2-norm in the first constraint by the Frobenius norm of the matrix. Indeed,  $\|M\|_2 \leq \|M\|_F$ , and the equality holds for rank-1 and zero matrices, so

$$\begin{cases} b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top V a_2 \geq \|D_1 V D_2^\top + p_1 z^\top D_2^\top + D_1 w p_2^\top + p_1 p_2^\top\|_F \\ b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top V a_2 \geq \|b_1 D_2 z + b_1 p_2 - D_2 V a_1 - a_1^\top w p_2\| \\ b_1 b_2 - b_1 a_2^\top z - b_2 a_1^\top w + a_1^\top V a_2 \geq \|b_2 D_1 w + b_2 p_1 - D_1 V a_2 - a_2^\top z p_1\| \end{cases}$$

are valid, yet looser, constraints.

We could even further tighten our approximations by applying RPT to  $z$  and to  $w$  separately. More specifically, we could multiply the linear and nonlinear constraints that define  $\mathcal{Z}$  with each other and express them in terms of  $z$ , an additional variable defined as  $Z = z z^\top$ . The same could be done for the constraints in  $\text{dom } h^*$ , and define  $W = w w^\top$ . By doing so, all the nonlinear substitutions are concisely defined as

$$\begin{pmatrix} W & V \\ V^\top & z \end{pmatrix} = \begin{pmatrix} w \\ z \end{pmatrix} \begin{pmatrix} w \\ z \end{pmatrix}^\top.$$

In a relaxation, the equality  $=$  is replaced by  $\succeq$  and, by using Schur complements, yields the semidefinite constraint:

$$\begin{pmatrix} W & V & w \\ V^\top & z & z \\ w^\top & z^\top & 1 \end{pmatrix} \succeq 0. \quad (11)$$

Hence, we can add this LMI to any of the safe approximations obtained by RPT  $\Theta_i$  and the resulting uncertainty set,  $\Theta_{LMI,i}$ , would provide a safe approximation at least as good as the one obtained from  $\Theta_i$ . However, whether the extra computational burden

outweighs the improvement of the approximation remains an open question, whose answer might depend on the specific problem. On a geometric optimization example in Sect. 5.2, for instance, we observed no numerical improvement from adding this LMI.

### 2.3 Step 3: Tractable robust counterpart

Finally, the safe approximation (10) is a robust constraint, linear in the uncertain parameters  $(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0)$ , and with a convex uncertainty set  $\Phi = \Theta_i$ ,  $i = 0, 1, 2, 3$ . Hence, following the approach of Ben-Tal et al. [4], one can derive its robust counterpart by computing the support function of  $\Phi$ ,  $\delta^*(\cdot | \Phi)$ . In particular, (10) is equivalent to

$$\delta^*((\mathbf{b}(\mathbf{x}), -1, \mathbf{0}, \mathbf{A}(\mathbf{x}), \mathbf{0}) | \Phi) \leq 0.$$

In our hierarchy, note that  $\Theta_{i+1}$  is obtained by imposing additional constraints to  $\Theta_i$ , hence is of the form  $\Theta_{i+1} = \Theta_i \cap \Phi_{i+1}$  for some convex set  $\Phi_{i+1}$ . Consequently, the support function of  $\Theta_{i+1}$  can be expressed as a function of the support functions of  $\Theta_i$  and  $\Phi_{i+1}$  [see [4, Lemma 6.4]].

### 2.4 Description of the approach for conic constraints

In this section, we describe our Reformulation-Perspectification-Technique for conic inequalities. We consider an uncertain conic constraint of the type

$$\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x}) \preceq_{\mathcal{K}} \mathbf{0}, \quad \forall \mathbf{z} \in \mathcal{Z}, \quad (12)$$

where  $\mathcal{K} \in \mathbb{R}^m$  is a closed convex cone with nonempty relative interior. Examples of these cones are the linear cone, the second-order cone, the semi-definite cone, the power cone, and the exponential cone. We refer the reader to MOSEK [16] for an extensive treatment of these five cones. The dual cone, referred to as  $\mathcal{K}^*$  of  $\mathcal{K}$  is defined as

$$\mathcal{K}^* = \left\{ \mathbf{w} \in \mathbb{R}^m : \mathbf{w}^\top \mathbf{u} \geq 0, \quad \forall \mathbf{u} \in \mathcal{K} \right\},$$

and is again a convex cone. Furthermore, we have  $\mathcal{K} = (\mathcal{K}^*)^*$ , because the cone is closed and convex. Using these properties, we derive for (12)

$$\begin{aligned} & \mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x}) \preceq_{\mathcal{K}} \mathbf{0}, \quad \forall \mathbf{z} \in \mathcal{Z} \\ \iff & \mathbf{w}^\top (\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x})) \leq 0, \quad \forall \mathbf{w} \in \mathcal{K}^* \quad \forall \mathbf{z} \in \mathcal{Z} \\ \iff & \max_{\mathbf{w} \in \mathcal{K}^*} \max_{\mathbf{z} \in \mathcal{Z}} \mathbf{w}^\top (\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x})) \leq 0 \\ \iff & \max_{\mathbf{w} \in \mathcal{K}^*} \max_{\mathbf{z} \in \mathcal{Z}} \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{w} \mathbf{z}^\top) + \mathbf{w}^\top \mathbf{b}(\mathbf{x}) \leq 0. \end{aligned}$$

**Table 2** Number of constraints defining the uncertainty sets of our hierarchy,  $\Theta_i$ ,  $i = 0, 1, 2$ 

Set	Linear constraints	Convex constraints
$\Theta_0$	$K_0 + J_0$	$K_1 + J_1 + 1$
$\Theta_1$	$K_0 + J_0 + K_0 J_0$	$K_1 + J_1 + 1 + K_0(J_1 + 1)$
$\Theta_2$	$K_0 + J_0 + K_0 J_0$	$K_1 + J_1 + 1 + K_0(J_1 + 1) + J_1 K_0$

We can now directly apply our RPT approach to the last inequality that is bilinear in the uncertain parameters  $\mathbf{w}$  and  $\mathbf{z}$ . Note that this conic case is a special case of (2) where we have  $h^*(\mathbf{w}) = 0$ , and  $\text{dom } h^* = \mathcal{K}^*$ .

## 2.5 Trade-off between approximation accuracy and computational complexity

Our RPT approach exhibits a hierarchy of increasingly tighter safe approximations. In this section, we discuss the trade-off between the quality of a safe approximation and its computational tractability.

In terms of computational tractability, Table 2 reports the number of linear and convex inequalities involved in each set of our hierarchy,  $\Theta_i$ ,  $i = 0, 1, 2$ . Observe that imposing constraints (8) for all  $j \in [J_0]$  is redundant with imposing (5) for all  $k \in [K_0]$ . Hence,  $\Theta_2$  requires no additional linear constraints compared with  $\Theta_1$ . Moving one level higher in the hierarchy requires adding a quadratic number of linear/convex constraints compared with the number of constraints defining  $\mathcal{Z}$  and  $\text{dom } h^*$ . Furthermore, our approach involves the perspective functions and conjugates of the functions defining  $\mathcal{Z}$  and  $\text{dom } h^*$ . If a function is conic representable in a cone  $\mathcal{K}$ , then its perspective (resp. conjugate) is conic representable in the same cone  $\mathcal{K}$  (resp. the dual cone  $\mathcal{K}^*$ ). On this aspect, our approach does not increase the computational complexity.

Regarding approximation guarantees, we note that if the description of  $\mathcal{Z}$  involves no linear constraints ( $K_0 = 0$ ), then  $\Theta_0 = \Theta_1$ , and  $\Theta_2$  would provide a substantial benefit. Alternatively, when  $\mathcal{Z}$  is polyhedral ( $J_0 = 0$ ),  $\Theta_1 = \Theta_2$  and  $\Theta_1$  provides all the benefit. Consequently, we intuit that the relative benefit of  $\Theta_2$  over  $\Theta_1$  depends on how binding the convex constraints in  $\mathcal{Z}$  are. For instance, if the description of  $\mathcal{Z}$  contains one convex constraint  $c(\mathbf{z}) \leq 0$  that is redundant with the linear constraints  $D\mathbf{z} \leq \mathbf{d}$ , then the constraints

$$(f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}) c \left( \frac{f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j}{f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w}} \right) \leq 0, \quad \forall j \in [J_0] \quad (13)$$

in  $\Theta_2$  are redundant with linear constraints in  $\Theta_1$ .

**Proof** As previously observed, imposing constraints (5) for all  $k \in [K_0]$  in  $\Theta_1$  is equivalent to imposing (8) for all  $j \in [J_0]$ . Hence, for any  $j \in [J_0]$ , for any  $(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta_1$ , if  $f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w} > 0$ ,  $\mathbf{y} := (f_j \mathbf{z} - \mathbf{F} \mathbf{V} \mathbf{e}_j) / (f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w})$  satisfies  $D\mathbf{y} \leq \mathbf{d}$ . Hence,  $c(\mathbf{y}) \leq 0$  and constraints (13) are satisfied. Note that the result holds if  $f_j - \mathbf{e}_j^\top \mathbf{F} \mathbf{w} = 0$ , by continuity.  $\square$

When  $\mathcal{Z}$  is polyhedral, we will show in Sect. 3.2 that RPT with  $\Theta_1$  can be viewed as approximating an adaptive robust optimization problem (with uncertainty set  $\text{dom } h^*$ ) with linear decision rules [as in [18, Theorem 2]]. In addition, if  $h$  is piecewise linear, then  $\text{dom } h^*$  is a simplex. Since linear decision rules are optimal for simplex uncertainty sets, as proved by Bertsimas and Goyal [7, Theorem 1] and generalized in Ben-Ameur et al. [1, Corollary 1],  $\Theta_1$  provides an exact reformulation of (1). In general, for polyhedral  $\mathcal{Z}$ , the quality of the approximation with  $\Theta_1$  will thus depend on the geometry of  $\text{dom } h^*$ , in particular its symmetry and “simplex dilation factor” [5]; see El Housni and Goyal [12, 13] and references therein for some average- and worst-case performance results of linear decision rules. Theoretical results to precisely quantify the benefit, in terms of approximation guarantee, of each level of the hierarchy constitutes an interesting future direction.

From a practical standpoint, when using our approach with the uncertainty set  $\Theta_1$ , one can check whether the additional constraints in  $\Theta_2$  are satisfied by the worst-case scenario for the current solution. If so, the current solution remains robust feasible for  $\Theta_2$  and  $\Theta_2$  will not provide any improvement. Alternatively, our approach provides both an upper and lower bound (see Sect. 4). Combined, they can be used to measure the approximation gap and decide whether to use  $\Theta_2$  instead.

### 3 Connection with previous approaches from the literature

In this section, we show that our RPT approach unifies many existing robust convex optimization techniques, and extends them in a significant way.

#### 3.1 Case when the uncertainty set lies in the non-negative orthant

First, we restrict our analysis to cases where the linear inequalities describing  $\mathcal{Z}$  in Assumption 1 are non-negativity constraints only. Formally, we make the following assumption:

**Assumption 3** The uncertainty set  $\mathcal{Z} = \{z \mid z \geq \mathbf{0}, c_k(z) \leq 0, k \in [K]\}$  is full dimensional and bounded, where  $c_k : \mathbb{R}^L \mapsto [-\infty, +\infty]$  is proper, closed and convex for each  $k \in [K]$ .

Note that the non-negativity assumption of  $\mathcal{Z}$  can always be satisfied by properly lifting or shifting the uncertainty set (see the detailed discussion in Sect. 5.1 and Appendix A). Under this more specialized assumption, we reformulate our RPT-based safe approximation with  $\Theta_1$  (Corollary 1) and discuss its connection with the approaches from Bertsimas and Sim [8] and Roos et al. [18]. We will later refer to Corollary 1 as the *perspectification approach* because it can also be derived without invoking the RPT but a perspectification lemma instead (as we do in Appendix A).

**Corollary 1** For  $\mathcal{Z}$  satisfying Assumption 3, the safe approximation obtained with RPT and  $\Theta_1$  is equivalent to

$$\sum_{\ell \in [L]} z_\ell h_\infty(A(x)e_\ell) + h(b(x)) \leq 0, \quad \forall z \in \mathcal{Z}. \quad (14)$$

**Proof** We need the following identity, for any  $u \geq 0$ ,

$$u h\left(\frac{\mathbf{y}}{u}\right) = \sup_{\mathbf{w} \in \text{dom } h^*} \left\{ \mathbf{y}^\top \mathbf{w} - u h^*(\mathbf{w}) \right\}.$$

Denoting between brackets the dual variables associated with constraints in the definition of  $\Theta_1$ ,

$$\Theta_1 = \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0) \left| \begin{array}{l} \mathbf{z} \in \mathcal{Z}, \\ \mathbf{w} \in \text{dom } h^*, h^*(\mathbf{w}) \leq w_0, \\ \mathbf{V} \mathbf{e}_\ell / z_\ell \in \text{dom } h^*, z_\ell h^*\left(\frac{\mathbf{V} \mathbf{e}_\ell}{z_\ell}\right) \leq v_{0\ell}, \ell \in [L], [u_\ell] \end{array} \right. \right\},$$

we have

$$\begin{aligned} & \sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0) \in \Theta_1} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \right\} \\ &= \sup_{\substack{\mathbf{z} \in \mathcal{Z} \\ w_0, v_0}} \sup_{\substack{\mathbf{w} \in \text{dom } h^* \\ \mathbf{V} \mathbf{e}_\ell / z_\ell \in \text{dom } h^*}} \inf_{u, u \geq 0} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 - u(h^*(\mathbf{w}) - w_0) \right. \\ & \quad \left. - \sum_{\ell \in [L]} u_\ell \left( z_\ell h^*\left(\frac{\mathbf{V} \mathbf{e}_\ell}{z_\ell}\right) - v_{0\ell} \right) \right\}. \end{aligned}$$

Let us observe that

$$\sup_{\mathbf{w} \in \text{dom } h^*} \mathbf{b}(\mathbf{x})^\top \mathbf{w} - u h^*(\mathbf{w}) = u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right),$$

and similarly for the terms involving  $\mathbf{V} \mathbf{e}_\ell / z_\ell$ ,  $\ell \in [L]$ , so that

$$\begin{aligned} & \sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, v_0) \in \Theta_1} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x})^\top \mathbf{V}) + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \right\} \\ &= \sup_{\mathbf{z} \in \mathcal{Z}} \inf_{u, u \geq 0} \sup_{w_0, v_0} \sum_{\ell} u_\ell z_\ell h\left(\frac{\mathbf{A}(\mathbf{x}) \mathbf{e}_\ell}{u_\ell}\right) + u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right) + \mathbf{u}^\top \mathbf{v}_0 + (u - 1)w_0 \\ &= \sup_{\mathbf{z} \in \mathcal{Z}} \sum_{\ell \in [L]} z_\ell h_\infty(\mathbf{A}(\mathbf{x}) \mathbf{e}_\ell) + h(\mathbf{b}(\mathbf{x})), \end{aligned}$$

where the last equality holds because  $\mathbf{u} = \mathbf{0}$  and  $u = 1$ . □

**Remark 1** Corollary 1 might be uninformative if the recession function of  $h$  is unbounded (e.g., if  $h$  is strongly convex). In this case, we propose an alternative to Corollary 1 with the additional requirement that  $h(\mathbf{0}) = 0$ . Indeed,  $h(\mathbf{0}) = 0$  implies  $w_0 \geq h^*(\mathbf{w}) \geq 0$ . In this case, we should define



$$\Theta_0 = \left\{ (w, w_0, z, V, v_0) \left| \begin{array}{l} z \in \mathcal{Z}, \\ w \in \text{dom } h^*, \\ w_0 \geq 0, h^*(w) \leq w_0, \\ v_0/w_0 \in \mathcal{Z} \end{array} \right. \right\},$$

in Proposition 2, where the membership  $v_0/w_0 \in \mathcal{Z}$  is enforced by multiplying each constraint in the definition of  $\mathcal{Z}$  by  $w_0 \geq 0$ . The resulting safe approximation is:

$$\begin{aligned} \exists (u, u) \in \mathbb{R}_+^{L+1}, \text{ s.t. } \sum_{\ell \in [L]} z_\ell u_\ell h\left(\frac{A(x)e_\ell}{u_\ell}\right) + u h\left(\frac{b(x)}{u}\right) \leq 0, \quad \forall z \in \mathcal{Z}, \\ z^\top u + u \leq 1, \quad \forall z \in \mathcal{Z}. \end{aligned}$$

The key benefit from Corollary 1 is that it preserves the uncertainty set. Namely, it safely approximates the robust convex constraint (1) by a robust constraint that is linear in the same uncertain parameter  $z$ . Consequently, one can also derive probabilistic guarantees for this safe approximation, after positing a generative distributional model for  $z$ . We derive such guarantees in Appendix A.

We now compare this formulation with the approaches from Bertsimas and Sim [8] and Roos et al. [18].

*For positively homogeneous functions  $h$ , we show that Corollary 1 recovers the approach of Bertsimas and Sim [8].*

Indeed, Bertsimas and Sim [8] construct a safe approximation for the robust constraint (1) in the special case where the function  $h(\cdot)$  is positively homogeneous, i.e.,  $h(\lambda y) = \lambda h(y)$  for any  $\lambda > 0$  and  $y$ . In this case,

$$h_\infty(y) = \lim_{t \rightarrow 0} t h(y/t) = h(y),$$

so Corollary 1 is equivalent to

$$\sum_{\ell \in [L]} z_\ell h(A(x)e_\ell) + h(b(x)) \leq 0, \quad \forall z \in \mathcal{Z},$$

which coincides with the approximation proposed by Bertsimas and Sim [8]. Accordingly, the RPT approach extends the one in Bertsimas and Sim [8] in multiple significant ways. First, Bertsimas and Sim [8] require  $h$  to be convex and homogeneous, hence sub-additive. They use sub-additivity to derive their safe approximation. In contrast, Corollary 1 applies to any convex function. In Appendix A, we show how Corollary 1 can also be obtained independently from the RPT approach, by using a generalization of sub-additivity to general convex functions called the perspectification lemma (Lemma 1). Second, the RPT approach can lead to tighter approximations when (a) some of the generic convex constraints  $c_k(z) \leq 0$  are also linear and can be used to define a smaller set  $\Theta_1$ , or when (b) a higher set in the hierarchy of approximations is used. Finally, besides the safe approximation, Bertsimas and Sim [8] also propose a non-convex lifting procedure to transform any norm-based set into a lifted uncertainty

set that lies in the non-negative orthant and satisfies Assumption 3. On that regard, our proposal is agnostic of the procedure used in practice to ensure Assumption 3 holds.

*We now show that Corollary 1 coincides with the safe approximation of (1) from Roos et al. [18] with static policies.*

Roos et al. [18] propose to first reformulate (1) as an adaptive robust *nonlinear* constraint, which they then approximate with static decision rules:

**Proposition 3** ([18], Theorem 7) *Under Assumption 3, the robust convex constraint (1) can be safely approximated by the constraints*

$$\begin{aligned} \exists \lambda \geq \mathbf{0}, \{\mathbf{u}^k\}_k, \text{ s.t. }, \sum_{k=1}^K \lambda_k c_k^* \left( \mathbf{u}^k / \lambda_k \right) + h(\mathbf{b}(\mathbf{x})) \leq 0, \\ h_\infty(\mathbf{A}(\mathbf{x})\mathbf{e}_\ell) \leq \sum_{k=1}^K \mathbf{u}_\ell^k, \forall \ell \in [L]. \end{aligned}$$

Proposition 3 is equivalent to Corollary 1 after observing that, for any  $\mathbf{y}$ ,

$$\sup_{z \geq 0: c_k(z) \leq 0, k \in [K]} \mathbf{y}^\top \mathbf{z} = \inf_{\substack{\mathbf{u}^k, k \in [K] \\ \lambda_k \geq 0, k \in [K]}} \sum_k \lambda_k c_k^* \left( \frac{\mathbf{u}^k}{\lambda_k} \right) \text{ s.t. } \mathbf{y} \leq \sum_{k \in [K]} \mathbf{u}^k.$$

Consequently, our RPT approach extends Roos et al. [18, Theorem 7] by providing tighter approximations when additional linear constraints are present or when using a higher set in the hierarchy.

### 3.2 Case when the uncertainty set is a polyhedron

*In this subsection, we show that the safe approximation obtained from  $\Theta_1$  for (1) with  $\mathcal{Z} = \{\mathbf{z} \mid \mathbf{D}\mathbf{z} \leq \mathbf{d}\}$ , where  $\mathbf{D} \in \mathbb{R}^{K_0 \times L}$ ,  $\mathbf{d} \in \mathbb{R}^{K_0}$ , coincides with that of Theorem 2 in Roos et al. [18].*

It follows from Proposition 2 that the convex robust constraint (1) can be safely approximated by

$$\sup_{(\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta_1} \left\{ \text{Tr}\{\mathbf{A}(\mathbf{x})^\top \mathbf{V}\} + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0 \right\} \leq 0,$$

where  $\Theta_1$  is equal to

$$\left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \left| \begin{array}{l} \mathbf{D}\mathbf{z} \leq \mathbf{d}, \\ h^*(\mathbf{w}) \leq w_0, \\ (d_k - \mathbf{e}_k^\top \mathbf{D}\mathbf{z}) h^* \left( \frac{\mathbf{w} d_k - \mathbf{V} \mathbf{D}^\top \mathbf{e}_k}{p_k - \mathbf{e}_k^\top \mathbf{D}\mathbf{z}} \right) \leq d_k w_0 - \mathbf{e}_k^\top \mathbf{D}\mathbf{v}_0, k \in [K_0] \end{array} \right. \right\}.$$

The corresponding tractable robust counterpart constitutes a safe approximation of (1), which coincides with the safe approximation of (1) proposed in Roos et al. [18, Theorem 2].

### 3.3 Connection to result for robust quadratic with an ellipsoid

It is well-known that robust quadratic inequalities with an ellipsoidal uncertainty set admit an exact SDP reformulation via the S-lemma [see [3]]. Alternatively, we show that the same reformulation can be obtained via RPT. To this end, consider the following robust quadratic inequality, which is a special case of the bilinear reformulation in Proposition 1 with  $\mathbf{w} = \mathbf{z}$ ,

$$\mathbf{z}^\top \mathbf{A}(\mathbf{x}) \mathbf{z} + \mathbf{b}(\mathbf{x})^\top \mathbf{z} \leq g(\mathbf{x}) \quad \forall \mathbf{z} \in \{\mathbf{z} \in \mathbb{R}^L : \mathbf{z}^\top \mathbf{D} \mathbf{z} + \mathbf{d}^\top \mathbf{z} \leq c\}, \quad (15)$$

which is equivalent to

$$\max_{(\mathbf{z}, \mathbf{Z}) \in \Theta_{ell}} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x}) \mathbf{Z}) + \mathbf{b}(\mathbf{x})^\top \mathbf{z} \right\} \leq g(\mathbf{x}),$$

where  $\Theta_{ell} = \{(\mathbf{z}, \mathbf{Z}) \in \mathbb{R}^L \times \mathbb{S}^L \mid \text{Tr}(\mathbf{D} \mathbf{Z}) + \mathbf{d}^\top \mathbf{z} \leq c, \mathbf{Z} = \mathbf{z} \mathbf{z}^\top\}$ . The set  $\Theta_{ell}$  is non-convex because of the nonlinear equality  $\mathbf{Z} = \mathbf{z} \mathbf{z}^\top$ , and the outer approximations of  $\Theta_{ell}$  proposed in Proposition 2 satisfy  $\Theta_{ell} \subseteq \Theta_{ell,0} = \Theta_{ell,1} = \Theta_{ell,2}$ . By relaxing the non-convex constraint  $\mathbf{Z} = \mathbf{z} \mathbf{z}^\top$  to  $\mathbf{Z} \succeq \mathbf{z} \mathbf{z}^\top$ , and then using Schur complement we obtain:

$$\Theta_{ell,3} = \left\{ (\mathbf{z}, \mathbf{Z}) \in \mathbb{R}^L \times \mathbb{S}^L \mid \text{Tr}(\mathbf{D} \mathbf{Z}) + \mathbf{d}^\top \mathbf{z} \leq c, \begin{pmatrix} \mathbf{Z} & \mathbf{z} \\ \mathbf{z}^\top & 1 \end{pmatrix} \succeq 0 \right\}.$$

The tractable robust counterpart of

$$\max_{(\mathbf{z}, \mathbf{Z}) \in \Theta_{ell,3}} \left\{ \text{Tr}(\mathbf{A}(\mathbf{x}) \mathbf{Z}) + \mathbf{b}(\mathbf{x})^\top \mathbf{z} \right\} \leq g(\mathbf{x})$$

coincides with the tractable robust counterpart of (15) obtained from using the S-lemma. Moreover, if the uncertainty set constitutes an intersection of ellipsoids, the obtained safe approximation for the robust quadratic inequality via RPT coincides with that from the approximate S-lemma [2]. In Zhen et al. [20], a similar relationship between the (approximate) S-lemma and RPT is established for quadratically constrained quadratic optimization.

## 4 Obtaining lower bounds

One simple way of relaxing constraint (1) is to consider a finite subset of scenarios  $\overline{\mathcal{Z}}$  sampled from the uncertainty set  $\mathcal{Z}$ , and the sampled version of (1) constitutes a finite set of convex inequalities:

$$h(A(x)z + b(x)) \leq 0, \forall z \in \mathcal{Z} \implies \max_{i \in [I]} h(A(x)z^{(i)} + b(x)) \leq 0 \\ \iff h(A(x)z^{(i)} + b(x)) \leq 0, \forall i \in [I], \quad (16)$$

where  $\{z^{(1)}, \dots, z^{(I)}\} = \overline{\mathcal{Z}} \subseteq \mathcal{Z}$ . Here, the point-wise maximum of convex functions in the sampled inequality can be reformulated into a finite set of convex inequalities (16), and provides a naive progressive approximation of (1).

The question that remains is how to choose the finite set of scenarios. One simple way to do this would be to consider all extreme points of  $\mathcal{Z}$ , and the sampled version (16) is then optimal. However, the number of extreme points of  $\mathcal{Z}$  could be large in general. Hadjiyiannis et al. [14] propose a way to obtain a small but effective finite set of scenarios for two-stage robust linear problems, which takes scenarios that are binding for the model solved with affine policies.

Alternatively, we can also adopt a dual perspective and sample scenarios from  $\text{dom } h^*$ . Consider the following form of the bilinear reformulation (2) with a fixed  $x$ :

$$\sup_{\substack{z \in \mathcal{Z} \\ w \in \text{dom } h^*}} \left\{ w^\top A(x)z + b(x)^\top w - h^*(w) \right\}. \quad (17)$$

The embedded maximization problem in (17) can be interpreted as a disjoint bilinear problem, that is, fixing either  $z$  or  $w$ , the embedded problem becomes a convex optimization problem. Hence, given a finite set  $\overline{\mathcal{W}} \subseteq \text{dom } h^*$ , the constraints

$$\sup_{z \in \mathcal{Z}} \left\{ w^\top A(x')z + b(x')^\top w - h^*(w) \right\} \leq 0, \forall w \in \overline{\mathcal{W}} \quad (18)$$

provide a progressive approximation of (1) as well.

Hence, we propose a hybrid approach to obtain valid lower bounds where we replace both  $\mathcal{Z}$  and  $\text{dom } h^*$  in (17) by finite subsets  $\overline{\mathcal{Z}}$  and  $\overline{\mathcal{W}}$  respectively. The idea here is adopted from Zhen et al. [21].

For a fixed  $x$ , we observe that the embedded problem in (17) becomes a convex optimization problem whenever  $z$  or  $w$  is fixed. By exploiting this observation, for any given  $z' \in \overline{\mathcal{Z}} \subseteq \mathcal{Z}$ , the maximizer

$$w' = \arg \max_{w \in \text{dom } h^*} w^\top A(x)z' + b(x)^\top w - h^*(w)$$

is a violating scenario if the corresponding optimal value is larger than 0, and  $w'$  can be used to update  $\overline{\mathcal{W}}$ , that is,  $\overline{\mathcal{W}} \leftarrow \overline{\mathcal{W}} \cup \{w'\}$ . Subsequently, fixing the obtained  $w'$  in (17), the maximizer

$$z' = \arg \max_{z \in \mathcal{Z}} w'^\top A(x)z + b(x)^\top w' - h^*(w) \quad (19)$$

is a violating scenario if the corresponding optimal value is larger than 0, and  $z'$  can be used to update  $\overline{\mathcal{Z}}$ , that is,  $\overline{\mathcal{Z}} \leftarrow \overline{\mathcal{Z}} \cup \{z'\}$ . We can repeat this iterative procedure till no improvement on the lower bound for (17) can be achieved. The enriched sets  $\overline{\mathcal{Z}}$  and

$\overline{\mathcal{W}}$  can then be used to improve  $\mathbf{x}$  by solving the optimization problem with both (16) with  $\overline{\mathcal{Z}}$  and (18) with  $\overline{\mathcal{W}}$ . The procedure can be repeated till no improvement can be achieved or the prescribed computational limit is met.

**Remark 2** One can improve the efficiency of the iterative procedure if  $A(\mathbf{x}) = A$  in (1), that is,  $A$  is independent of  $\mathbf{x}$ . In this case, for a given  $\mathbf{w}'$ , de Ruiter et al. [19] have shown that the maximizer  $\mathbf{z}'$  in (19) dominates  $\mathbf{w}'$ , that is, the feasible region of  $\mathbf{x}$  in (16) with  $\overline{\mathcal{Z}}' = \{\mathbf{z}'\}$  is larger or equal to the one in (18) with  $\overline{\mathcal{W}}' = \{\mathbf{w}'\}$ . Therefore, the inequalities involving  $\mathbf{w}'$  is redundant, and can be omitted throughout the procedure to improve the efficiency of this iterative procedure.

## 5 Computational results

In this section, we assess the numerical benefit from our RPT hierarchy on two examples, a robust control and geometric optimization setting, respectively.

### 5.1 Constrained linear-quadratic control

In this section, we consider an example from stochastic linear-quadratic control as in Bertsimas and Brown [6]. They consider a discrete-time stochastic system of the form:

$$\mathbf{y}_{k+1} = \mathbf{A}_k \mathbf{y}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{C}_k \mathbf{z}_k, \quad k = 0, \dots, N-1,$$

where  $\mathbf{y}_k$ ,  $\mathbf{u}_k$ , and  $\mathbf{z}_k$  are the state, control, and disturbance vectors respectively. The objective is to control the system (find control vectors  $\mathbf{u}_k$ ) in order to minimize the cost function

$$J(\mathbf{y}_0, \mathbf{u}, \mathbf{z}) = \sum_{k=1}^N (\mathbf{y}_k^\top \mathbf{Q}_k \mathbf{y}_k + 2\mathbf{q}_k^\top \mathbf{y}_k) + \sum_{k=0}^{N-1} (\mathbf{u}_k^\top \mathbf{R}_k \mathbf{u}_k + 2\mathbf{r}_k^\top \mathbf{u}_k),$$

under some uncertainty on the disturbance vectors  $\mathbf{z}_k$ . After algebraic manipulations and given a robust description of the uncertainty, the problem can be cast into

$$\min_{\mathbf{x}, t} t \quad \text{s.t.} \quad \|\mathbf{x}\|_2^2 + 2(\mathbf{F}^\top \mathbf{x} + \mathbf{h})^\top \mathbf{z} + \mathbf{z}^\top \mathbf{C} \mathbf{z} \leq t, \quad \forall \mathbf{z} \in \mathcal{Z},$$

where  $\mathcal{Z} = \{\mathbf{z} : \|\mathbf{z}\|_2 \leq \gamma\}$ , and properly define vector  $\mathbf{h}$  and matrices  $\mathbf{F}, \mathbf{C}$  ( $\mathbf{C} \succeq \mathbf{0}$ ). Note that  $\mathbf{x}$  can be subject to other deterministic constraints. For instance, imposing  $\mathbf{u}_k \geq \mathbf{0}$  yields linear constraints on  $\mathbf{x}$ . We refer the reader to Bertsimas and Brown [6] and references therein for problem motivation and details on the derivations.

Hence, we focus our attention to the robust convex constraint

$$2(\mathbf{F}^\top \mathbf{x} + \mathbf{h})^\top \mathbf{z} + \mathbf{z}^\top \mathbf{C} \mathbf{z} \leq t - \|\mathbf{x}\|_2^2, \quad \forall \mathbf{z} \in \mathcal{Z}. \quad (20)$$

Note that  $\mathcal{Z}$  is described using a single quadratic constraint. However, our approximations partly rely on the presence of linear constraints in the definition of the uncertainty

set. Hence, one could consider a shifted uncertainty set instead, obtained by introducing redundant linear inequalities such as  $\underline{z} \leq z \leq \bar{z}$  with  $\bar{z} = \gamma e$  and  $\underline{z} = -\gamma e$ , or a non-convex lifting of  $\mathcal{Z}$  as in Bertsimas and Sim [8]. In this section, we compare these modeling alternatives with three objectives in mind: (a) illustrate the benefit of our general RPT approach outlined in Sect. 2 over the perspectification approach from Sect. 3.1 (and Appendix A), which requires more stringent assumptions on the uncertainty set, (b) measure the relative benefit from shifting or lifting the uncertainty set, and (c) compare and assess the overall strength of our proposed safe approximations.

It is well-known that the robust constraint (20) is equivalent to a semidefinite constraint, which we will use as a benchmark to measure the suboptimality of our approach. In our numerical experiments, as in Bertsimas and Brown [6], we take  $y_0 = -1$ ,  $A_k = B_k = C_k = 1$ ,  $Q_k = R_k = \beta^k$  with  $\beta \in (0, 1)$ , and  $q_k = r_k = 0$ .

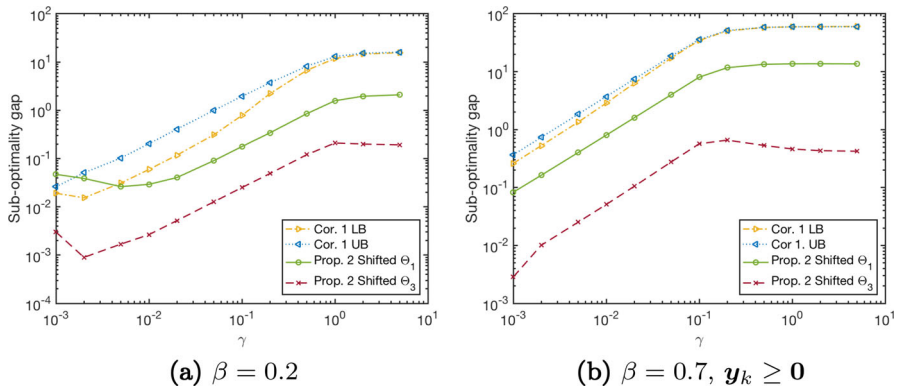
### Benefit from RPT over Perspectification

As we derived in Sect. 3.1, applying RPT up to  $\Theta_1$  in Proposition 2 leads to the safe approximation presented in Corollary 1. In addition, here, there are no linear constraints in the description of  $\text{dom } h^*$ , so  $\Theta_1 = \Theta_2$  and the benefit from using RPT (Proposition 2) over perspectification (Corollary 1) might be unclear. Yet, as discussed in Sect. 2.2, RPT is more flexible and can allow for an arbitrary number of linear constraints, while perspectification can only account for non-negativity constraints on  $z$ . We illustrate this point on the present example.

To derive meaningful safe approximations, we need to enrich the definition of  $\mathcal{Z}$  with linear constraints. A generic method for doing so is shifting the uncertainty set. For instance, any  $z : \|z\|_2 \leq \gamma$  satisfies  $-\gamma e \leq z \leq \gamma e$ . Accordingly, we can:

- Consider the shifted uncertain vector  $\tilde{z} = \gamma e + z$ ,  $z \in \mathcal{Z}$ , so that  $\tilde{z} \geq 0$  and apply Corollary 1. We will refer to this method as [Cor. 1 LB];
- Consider the shifted uncertainty set  $\gamma e - \mathcal{Z} \subseteq \mathbb{R}_+^I$  and apply Corollary 1 [Cor. 1 UB];
- Add the  $2L$  linear inequalities  $-\gamma e \leq z \leq \gamma e$  to the definition of  $\mathcal{Z}$  and apply RPT with  $\Theta_1$  [Prop. 2, Shifted  $\Theta_1$ ].

Derivations of the resulting three safe approximations are provided in Appendix B. Figure 1 displays the sub-optimality gap of each alternative, compared with the exact SDP reformulation, as  $\gamma$  increases. [Prop. 2, Lifted  $\Theta_1$ ] provides a substantial improvement over both [Cor. 1 LB] and [Cor. 1 UB], hence illustrating the benefit from accounting for both lower and upper bounds on  $z$  when safely approximating the robust constraint. Corollary 1, however, is unable to leverage such information, for it requires exactly and only  $L$  non-negativity constraints. In addition, applying RPT opens the door to a hierarchy of approximations. For instance, by multiplying the two norm constraints in the definition of  $\mathcal{Z}$  and  $\text{dom } h^*$  respectively, one can consider the third level of approximation  $\Theta_3$  which further reduces the optimality gap (see performance of [Prop. 2, Lifted  $\Theta_3$ ] in Fig. 1).



**Fig. 1** Computational results on constrained linear-quadratic control. Evolution of the sub-optimality gap of the safe approximation as  $\gamma$  increases for a constrained linear-quadratic control examples with  $N = 20$  periods and shifted uncertainty sets

### Benefit from lifting over shifting

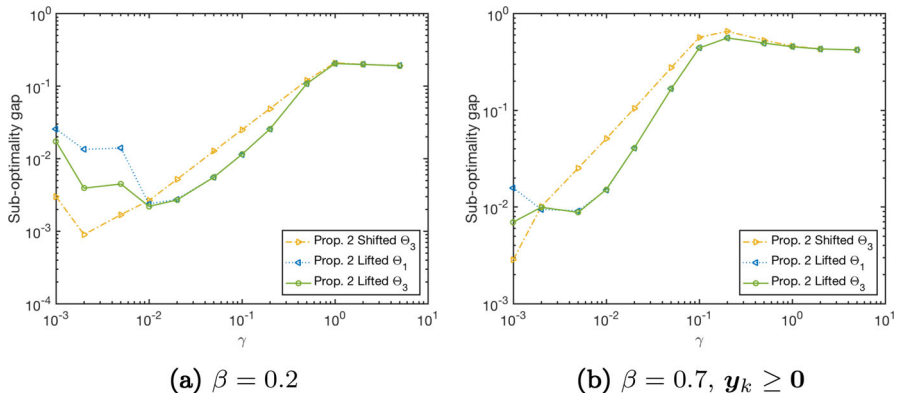
Another option to add linear constraints into the definition of the uncertainty set is to decompose  $z$  into  $z = z^+ - z^-$  and consider the lifted uncertainty set  $\{(z^+, z^-) : z^+, z^- \geq 0, \|z^+ + z^-\|_2 \leq \gamma\}$ . Observe that this kind of non-convex lifting cannot be applied to any uncertainty set, for the constraint  $\|z\|_2 \leq \gamma$  is replaced by  $\|z^+ + z^-\|_2 \leq \gamma$  - instead of  $\|z^+ - z^-\|_2 \leq \gamma$  - and constitutes a valid lifting of  $\mathcal{Z}$  for norm-sets only [8]. Yet, as demonstrated on Fig. 2, such non-convex lifting of the uncertainty set can lead to tighter safe approximations (see [Prop. 2, Lifted  $\Theta_1$ ] and [Prop. 2, Lifted  $\Theta_3$ ]) than shifting, especially for low to moderate values of  $\gamma$ . We remark that the first level of approximation with the lifted uncertainty set [Prop. 2, Lifted  $\Theta_1$ ] - which is equivalent to the approach outlined in Bertsimas and Sim [8] - competes and even surpasses the third level of approximation on the shifted one. However, we observe that the difference between shifting and lifting shrinks as  $\gamma$  increases. Also, we should emphasize the fact that shifting is a very generic procedure that can be applied to any bounded uncertainty set, without further assumption on its structure.

### Overall performance

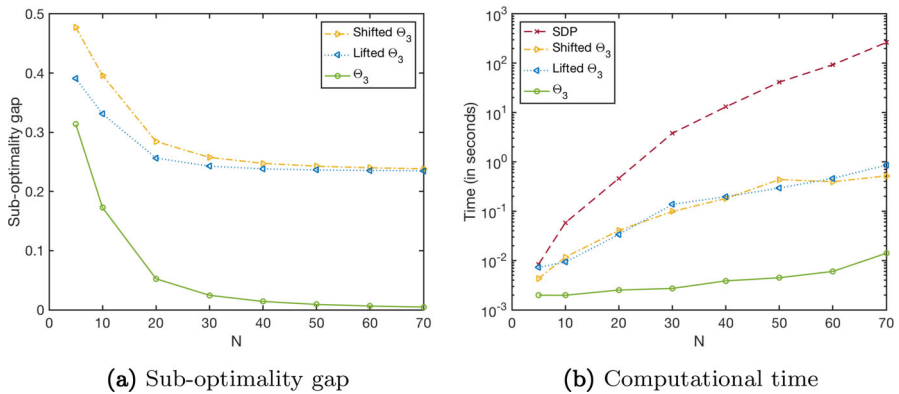
Figure 3 compares the performance of the SDP reformulation with safe approximations obtained from  $\Theta_3$  in Proposition 2, with the original uncertainty set  $\{z : \|z\|_2 \leq \gamma\}$ , its shifted and lifted versions respectively. In short, we observe that our approach successfully provides near-optimal solution to (within 5% optimality) at a fraction of the cost of solving the exact SDP reformulation.

## 5.2 Robust geometric optimization

We evaluate the performance of our proposed approach on geometric optimization instances that are randomly as in Hsiung et al. [15]. In particular, we consider geometric



**Fig. 2** Computational results on constrained linear-quadratic control. Relative performance of shifting vs. lifting the uncertainty set as  $\gamma$  increases for constrained linear-quadratic control examples with  $N = 20$  periods. Performance is measured in terms of sub-optimality gap with respect to the SDP reformulation



**Fig. 3** Computational results on constrained linear-quadratic control. Numerical behavior of three level-3 safe approximations derived with RPT (Proposition 2) as the number of time periods  $N$  increases, compared with the exact SDP reformulation of the robust constraint. For this experiment, we set  $\beta = 1.0$  and impose  $\mathbf{y}_k \geq \mathbf{0}$

optimization problems with a linear objective, and a system of two-term log-sum-exp robust inequalities

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \log \left( e^{(-1 + \mathbf{B}_i^{(1)} \mathbf{z})^\top \mathbf{x}} + e^{(-1 + \mathbf{B}_i^{(2)} \mathbf{z})^\top \mathbf{x}} \right) \leq 0, \quad \forall \mathbf{z} \in \mathcal{Z}, \quad i \in [I], \end{aligned} \quad (21)$$

where  $\mathbf{c} = \mathbf{1} \in \mathbb{R}^{n_x}$  is the all ones vector, and  $\mathbf{B}_i^{(1)}, \mathbf{B}_i^{(2)} \in \mathbb{R}^{n_x \times L}$  are randomly generated sparse matrices with sparsity density 0.1 whose nonzero elements are uniformly distributed on the interval  $[-1, 1]$ . The uncertainty set is assumed to be an intersection of a hypercube and a ball, that is,



$$\mathcal{Z} = \left\{ \mathbf{z} \in \mathbb{R}^L \mid \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_2 \leq \gamma \right\}.$$

We consider a set of instances with  $n_x = I = 100$  and  $L \in \{6, 8, \dots, 20\}$ . We select a  $\gamma = \sqrt{\frac{2^L \Gamma(n/2+1)}{\pi^{L/2}}}$ , where  $\Gamma$  denotes the gamma function, that ensures the volume of the ball  $\{\mathbf{z} \mid \|\mathbf{z}\|_2 \leq \gamma\}$  coincides with the volume of the hypercube  $\{\mathbf{z} \mid \|\mathbf{z}\|_\infty \leq 1\}$ . All the reported numerical results are from the average of 10 randomly generated instances.

Consider the bilinear reformulation of the  $i$ -th constraint of (21)

$$\sup_{(\mathbf{w}, \mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0) \in \Theta} \left\{ \left( -\mathbf{1}w_1 + \mathbf{B}_i^{(1)} \mathbf{v}_1 \right)^\top \mathbf{x} + \left( -\mathbf{1}w_2 + \mathbf{B}_i^{(2)} \mathbf{v}_2 \right)^\top \mathbf{x} - w_0 \right\}$$

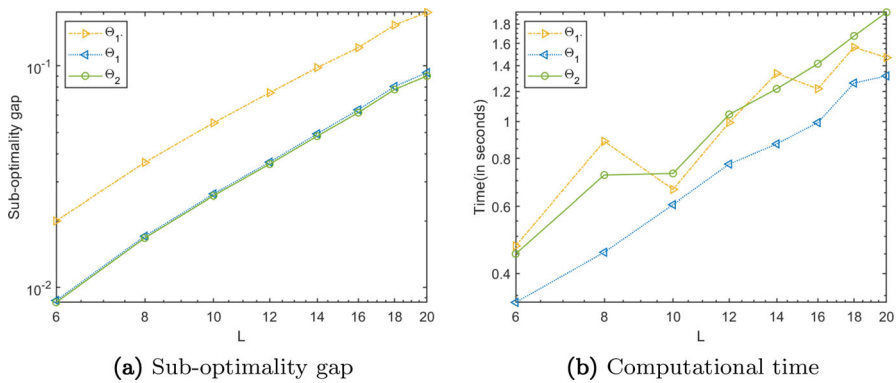
with the non-convex uncertainty set

$$\Theta = \left\{ (\mathbf{w}, \mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0) \mid \begin{array}{ll} \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_2 \leq \gamma, & (\mathbf{z} \in \mathcal{Z}) \\ w_1, w_2 \geq 0, w_1 + w_2 = 1, & (\mathbf{w} \in \text{dom } h^*) \\ w_1 \log w_1 + w_2 \log w_2 \leq w_0, & (\text{epigraph}) \\ \mathbf{v}_1 = w_1 \mathbf{z}, \mathbf{v}_2 = w_2 \mathbf{z}, \\ \mathbf{v}_0 = w_0 \mathbf{z}, \end{array} \right\},$$

which is safely approximated by four convex sets that satisfy  $\Theta \subseteq \Theta_{LMI,2} \subseteq \Theta_2 \subseteq \Theta_1 \subseteq \Theta_{1'}$ , where  $\Theta_1$  and  $\Theta_2$  are obtained from applying Proposition 2 to  $\Theta$ , while  $\Theta_{LMI,2}$  is obtained by enriching  $\Theta_2$  with the LMI in (11). In addition, we also compare the safe approximations from  $\Theta_1$ ,  $\Theta_2$  and  $\Theta_{LMI,2}$  with  $\Theta_{1'}$ , where  $\Theta_{1'}$  is the corresponding  $\Theta_1$  if  $\mathcal{Z}_{box} = \{\mathbf{z} \mid \|\mathbf{z}\|_\infty \leq 1\}$  is considered instead of  $\mathcal{Z}$ . Note that the obtained safe approximation from  $\Theta_{1'}$  coincides with that of Roos et al. [18] as discussed in Sect. 3.2. We refer to Appendix C.1 for a detailed representation of  $\Theta_1$ ,  $\Theta_{1'}$ ,  $\Theta_2$  and  $\Theta_{LMI,2}$ . We resort to comparing our solutions' objective value to a lower bound. To this end, we compute a lower bound by using the optimal solution to the tightest obtained safe approximation to find potentially critical scenarios in the uncertainty set. The lower bound is then constructed by solving a model that only safeguards for this finite set of critical scenarios; see more detail in Sect. 4.

## Numerical performance

From Fig. 4 we observe that, while our proposed safe approximations from  $\Theta_1$  and  $\Theta_2$  requires similar computational effort as that from  $\Theta_{1'}$ , the average optimality gaps from considering  $\Theta_1$  and  $\Theta_2$  are consistently smaller than that from considering  $\Theta_{1'}$ . For all considered instances, we observe that despite the significant additional computational effort, the optimality gaps obtained from considering  $\Theta_{LMI,2}$  coincide with the optimality gaps from considering  $\Theta_2$ . Therefore, we do not report the computation time for the safe approximations from  $\Theta_{LMI,2}$ . Moreover, we also consider the shifted uncertainty set, and observe that the obtained optimality gaps coincide with the optimality gaps when the original uncertainty set is considered.



**Fig. 4** Computational results on robust geometric optimization. Numerical behavior of the safe approximations from  $\Theta_1^+$ ,  $\Theta_1$  and  $\Theta_2$  as the size of the instance, that is,  $L \in \{6, 8, \dots, 20\}$ , increases

### Benefit from lifting

Alternatively, we consider the following lifted uncertainty set of  $\mathcal{Z}$ .

$$\mathcal{Z}^+ = \left\{ (z^+, z^-) \in \mathbb{R}_+^L \times \mathbb{R}_+^L \mid \|z^+ + z^-\|_\infty \leq 1, \|z^+ + z^-\|_2 \leq \gamma \right\}.$$

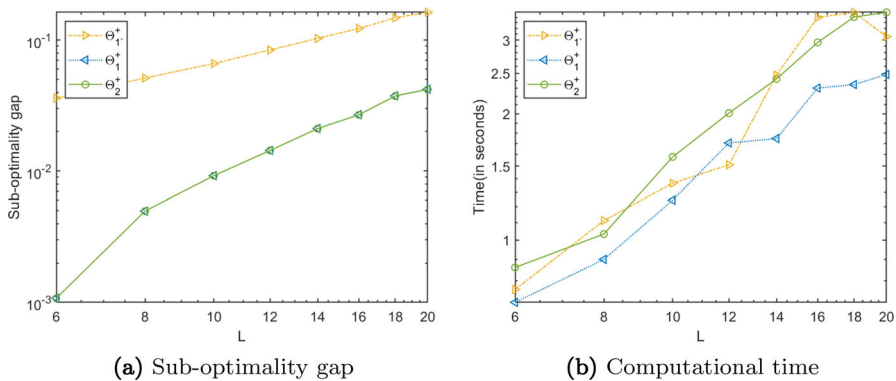
This non-convex projection is first proposed in Bertsimas and Sim [8] and then extended to solve adaptive robust linear problems in Chen and Zhang [11]. By substituting  $z$  with  $z^+ - z^-$ , the bilinear reformulation becomes

$$\sup_{(w, z^\pm, v_1^\pm) \in \Theta^+} (-w_1 \mathbf{1} + B_i^{(1)}[v_1^+ - v_1^-])^\top x + (-w_2 \mathbf{1} + B_i^{(2)}[v_2^+ - v_2^-])^\top x - w_0$$

with the lifted non-convex uncertainty set

$$\Theta^+ = \left\{ \begin{pmatrix} w, z^+, z^-, v_1^+, v_1^- \\ v_2^+, v_2^-, v_0^+, v_0^- \end{pmatrix} \mid \begin{array}{l} \|z^+ + z^-\|_\infty \leq 1, \\ \|z^+ + z^-\|_2 \leq \gamma, \quad ((z^+, z^-) \in \mathcal{Z}^+) \\ z^+, z^- \in \mathbb{R}_+^L \\ w_1, w_2 \geq 0, w_1 + w_2 = 1, \quad (w \in \text{dom} h^*) \\ w_1 \log w_1 + w_2 \log w_2 \leq w_0, \quad (\text{epigraph}) \\ v_1^+ = w_1 z^+, v_1^- = w_1 z^-, \\ v_2^+ = w_2 z^+, v_2^- = w_2 z^-, \\ v_0^+ = w_0 z^+, v_0^- = w_0 z^-, \end{array} \right\},$$

which is safely approximated by four convex sets that satisfy  $\Theta^+ \subseteq \Theta_2^+ \subseteq \Theta_1^+ \subseteq \Theta_1^+$  and  $\Theta^+ \subseteq \Theta_2^+ \subseteq \Theta_1^+ \subseteq \Theta_{BS}$ , where  $\Theta_1^+$  and  $\Theta_2^+$  obtained from applying Proposition 2 to  $\Theta^+$ , and  $\Theta_1^+$  is the corresponding  $\Theta_1^+$  if  $\mathcal{Z}_{box}^+ = \{(z^+, z^-) \in \mathbb{R}_+^L \times \mathbb{R}_+^L \mid \|z^+ + z^-\|_\infty \leq 1\}$  is considered instead of  $\mathcal{Z}^+$ . We define  $\Theta_{BS}$  in such a way that the obtained safe approximation coincides with that of the extended approach



**Fig. 5** Computational results on robust geometric optimization. Numerical behavior of the safe approximations from  $\Theta_{1'}^+$ ,  $\Theta_1^+$  and  $\Theta_2^+$  as the size of the instance, that is,  $L \in \{6, 8, \dots, 20\}$ , increases

of Bertsimas and Sim [8]; see Sects. 3.1 and Appendix A. We refer to Appendix C.2 for a detailed representation of  $\Theta_{BS}$ ,  $\Theta_1^+$ ,  $\Theta_{1'}^+$  and  $\Theta_2^+$ . We do not consider  $\Theta_{LMI,2}^+$ , that is,  $\Theta_2^+$  with addition LMI similarly as in (11), because we already observe that  $\Theta_{LMI,2}$  does not improve the approximation from  $\Theta_2$  for the problem we consider here.

### Numerical performance with lifting

From Fig. 5 we observe that, while our proposed safe approximations from  $\Theta_1^+$  and  $\Theta_2^+$  require similar computational effort as that from  $\Theta_{BS}$ , the average optimality gaps from considering  $\Theta_1^+$  and  $\Theta_2^+$  are consistently smaller than that from considering  $\Theta_{BS}$ . For all considered instances, we observe that the optimality gaps obtained from considering  $\Theta_{1'}^+$  coincide with the optimality gaps from considering  $\Theta_1$  (i.e., without lifting). Therefore, we do not report the computation time for the safe approximations from  $\Theta_{1'}^+$ .

### Overall performance

Firstly, the average optimality gaps from  $\Theta_{1'}$  (which has the same performance as Roos et al. [18]; see Sect. 3.2) and  $\Theta_{BS}$  (which has the same performance as the extended approach of Bertsimas and Sim [8]; see Sects. 3.1 and Appendix A) are very close to each other, while the computational effort required for computing the safe approximation from  $\Theta_{BS}$  is almost twice as much as that from  $\Theta_{1'}$ . Our proposed safe approximations from  $\Theta_1$ ,  $\Theta_2$ ,  $\Theta_1^+$  and  $\Theta_2^+$  consistently outperform that from considering  $\Theta_{1'}$  and  $\Theta_{BS}$ . Although the computational effort required for computing the safe approximations from  $\Theta_1^+$  and  $\Theta_2^+$  is more than twice of that from  $\Theta_{1'}$  and  $\Theta_{BS}$ , the average optimality gaps from  $\Theta_1^+$  and  $\Theta_2^+$  are half of the ones from  $\Theta_{1'}$  and  $\Theta_{BS}$ .

## 6 Conclusion

We propose a hierarchical safe approximation scheme for (1). Via numerical experiments, we demonstrate that our approach either coincides with or improves upon the existing approaches of Bertsimas and Sim [8] and Roos et al. [18]. Furthermore, our approach not only provides a trade-off between the solution quality and the computational effort, but also allows a direct integration with existing safe approximation schemes, e.g., the lifting technique proposed in Bertsimas and Sim [8] and Chen and Zhang [11], which is used to further improve the obtained safe approximations.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10107-022-01881-w>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A The perspectification approach

In this section, we propose an extension of the approach proposed by Bertsimas and Sim [8] for general convex constraints and general convex uncertainty sets. Although this approach is dominated by our main RPT-based approach in Sect. 2, it has the advantage that probability guarantees can be derived for the solutions obtained. While Bertsimas and Sim [8] requires the function  $h$  to be sub-additive, our generalization is based on the ‘perspectification’ of the function  $h(\cdot)$ .

As in Bertsimas and Sim [8], we make the additional assumption that the uncertainty set lies in the non-negative orthant, as stated in Assumption 3. The non-negativity assumption of  $\mathcal{Z}$  holds without loss of generality and can always be satisfied by properly lifting or shifting the uncertainty set. Indeed, if  $\mathcal{Z} \not\subseteq \mathbb{R}_+^L$ , one can decompose  $z$  into  $z = z^+ - z^-$  with  $z^+, z^- \geq \mathbf{0}$ . Then,

$$A(x)z + b(x) = A(x)z^+ - A(x)z^- + b(x) = \tilde{A}(x) \begin{bmatrix} z^+ \\ z^- \end{bmatrix} + b(x),$$

where  $\tilde{A}(x) = [A(x), -A(x)]$  is linear in  $x$  and our analysis applies to the lifted uncertainty set

$$\mathcal{Z}' = \left\{ (z^+, z^-) \in \mathbb{R}_+^L \times \mathbb{R}_+^L : (z^+ - z^-) \in \mathcal{Z} \right\}.$$

However, the lifted uncertainty set  $\mathcal{Z}'$  defined above might be unbounded. As a result, Bertsimas and Sim [8] and Chen and Zhang [11] propose to incorpo-

rate a non-convex lifting for norm-based uncertainty sets. For instance, if  $\tilde{\mathcal{Z}} = \{z \in \mathbb{R}^L : z \in \mathcal{Z}, \|z\|_\infty \leq 1, \|z\|_2 \leq \gamma\}$ , Bertsimas and Sim [8] consider the following lifted uncertainty set:

$$\mathcal{Z}^+ = \left\{ (z^+, z^-) \in \mathbb{R}_+^L \times \mathbb{R}_+^L : \begin{array}{l} \|z^+ + z^-\|_\infty \leq 1, \\ \|z^+ + z^-\|_2 \leq \gamma \end{array} \right\}.$$

Alternatively, we can shift the uncertainty set by some constant factor  $z_0$  such that  $z + z_0 \in z_0 + \mathcal{Z} \subseteq \mathbb{R}_+^L$ . Hence,

$$A(x)z + b(x) = A(x)(z + z_0) \underbrace{- A(x)z_0 + b(x)}_{=: \tilde{b}(x) \text{ affine in } x}.$$

We numerically discussed the implications of these two alternatives on the numerical behavior of our proposed approximations in Sect. 5.1.

As previously discussed in Sect. 3.1, the perspectification lemma we present in this section offers an alternative derivation of Corollary 1, i.e., the RPT-based approximation with  $\Theta_1$  under Assumption 3. In addition, for this approach, we derive probabilistic guarantees, i.e., bounds on the probability of constraint violation, under the assumption that the uncertain parameter has sub-Gaussian tails.

### A.1 Safe approximations via perspectification

The perspectification approach is based on the following perspectification lemma [19]:

**Lemma 1** (Perspectification of Convex Functions) *If the function  $g : \mathbb{R}^m \mapsto [-\infty, +\infty]$  is proper, closed and convex, then for any  $y_1, \dots, y_m \in \mathbb{R}^m$  and non-negative weights  $\alpha_1, \dots, \alpha_m \geq 0$ , we have:*

$$g\left(\sum_{i=1}^m \alpha_i y_i\right) \leq \sum_{i=1}^m \alpha_i t_i g\left(\frac{y_i}{t_i}\right), \quad (22)$$

for any  $t \geq 0$  satisfying  $\sum_{i=1}^m \alpha_i t_i = 1$ .

**Remark 3** If  $g$  is positively homogeneous, i.e., for any vector  $y \in \mathbb{R}^m$  and any positive scalar  $\lambda > 0$ ,  $g(\lambda y) = \lambda g(y)$ , then (22) reduces to

$$g\left(\sum_{i=1}^m \alpha_i y_i\right) \leq \sum_{i=1}^m \alpha_i g(y_i),$$

and the additional variables  $t_i$  in (22) need not to be introduced. In this case, Lemma 1 simply states that  $g$  is sub-additive.

**Remark 4** The terms in the right-hand side of Lemma 1 should be interpreted as  $\alpha_i$  times the perspective function of  $g$  at  $(t_i, y_i)$ . In particular, it is defined when  $t_i = 0$  and equal to  $\alpha_i g_\infty(y_i)$  in this case.

We use this lemma to develop a *safe approximation* of (1), i.e., a sufficient condition for  $\mathbf{x}$  to satisfy (1).

**Proposition 4** (Safe Approximation via Perspectification) *Under Assumption 3, a safe approximation of (1) is:*

$$\forall \mathbf{z} \in \mathcal{Z}, \exists (\mathbf{u}, u) \in \mathbb{R}_+^{L+1}, \text{ s.t. } \sum_{\ell \in [L]} z_\ell u_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{u_\ell}\right) + u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right) \leq 0, \quad (23a)$$

$$\sum_{i=1}^L z_i u_i + u = 1, \quad (23b)$$

where  $\mathbf{A}(\mathbf{x})\mathbf{e}_\ell$  selects the  $\ell$ -th column of the matrix  $\mathbf{A}(\mathbf{x})$  for each  $\ell \in [L]$ .

**Proof** Consider  $\mathbf{z} \in \mathcal{Z}$ . If there exists a vector  $(\mathbf{u}, u) \in \mathbb{R}_+^{L+1}$  satisfying (23a)–(23b),  $(\mathbf{u}, u)$  satisfies the assumption of Lemma 1, (23b). Hence,

$$\begin{aligned} h(\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x})) &= h\left(\sum_{\ell \in [L]} z_\ell \mathbf{A}(\mathbf{x})\mathbf{e}_\ell + \mathbf{b}(\mathbf{x})\right) \\ &\leq \sum_{\ell \in [L]} z_\ell u_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{u_\ell}\right) + u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right). \end{aligned}$$

Finally, the right-hand side is lower than 0 from (23a) so (1) holds.  $\square$

Following Remark 4, the terms  $u_\ell h(\mathbf{A}(\mathbf{x})\mathbf{e}_\ell/u_\ell)$  in (23a) should be interpreted as  $\text{persp}_h(u_\ell, \mathbf{A}(\mathbf{x})\mathbf{e}_\ell)$ . In particular, for  $u_\ell = 0$ ,  $u_\ell h(\mathbf{A}(\mathbf{x})\mathbf{e}_\ell/u_\ell) = h_\infty(\mathbf{A}(\mathbf{x})\mathbf{e}_\ell)$ . Proposition 4 approximates the robust convex constraint (1) via a set of adaptive robust constraints that depend linearly in the uncertainty  $\mathbf{z}$ . Note, however, that even in the fully adaptive case where  $(\mathbf{u}, u)$  can depend arbitrarily on the uncertain parameter  $\mathbf{z}$ , Proposition 4 is only a valid approximation of the robust constraint (1), not an equivalent reformulation.

**Remark 5** The safe approximation in Proposition 4 can be concisely written as

$$\sup_{\mathbf{z} \in \mathcal{Z}} \inf_{\substack{(\mathbf{u}, u) \in \mathbb{R}_+^{L+1} \\ \mathbf{z}^\top \mathbf{u} + u = 1}} \sum_{\ell \in [L]} z_\ell u_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{u_\ell}\right) + u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right) \leq 0.$$

Expressing  $h(\cdot)$  as the conjugate of its conjugate, we have

$$\begin{aligned} z_\ell u_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{u_\ell}\right) &= \sup_{\mathbf{v}_\ell: \mathbf{v}_\ell / z_\ell \in \text{dom } h^*} \left\{ \mathbf{e}_\ell^\top \mathbf{A}(\mathbf{x})^\top \mathbf{v}_\ell - z_\ell u_\ell h^*\left(\frac{\mathbf{v}_\ell}{z_\ell}\right) \right\}, \\ u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right) &= \sup_{\mathbf{w} \in \text{dom } h^*} \left\{ \mathbf{b}(\mathbf{x})^\top \mathbf{w} - u h^*(\mathbf{w}) \right\}. \end{aligned}$$

Finally, for a fixed  $\mathbf{z} \in \mathcal{Z}$ , we can interchange the inf and sup operators, take the dual with respect to  $(\mathbf{u}, u)$ , and reformulate Proposition 4 as

$$\sup_{\mathbf{z} \in \mathcal{Z}} \sup_{\substack{\mathbf{w} \in \text{dom } h^* \\ \frac{\mathbf{v}_\ell}{z_\ell} \in \text{dom } h^*}} \sup_{\substack{w_0: h^*(\mathbf{w}) \leq w_0 \\ z_\ell h^*(\frac{\mathbf{v}_\ell}{z_\ell}) \leq z_\ell w_0}} \sum_{\ell \in [L]} \mathbf{e}_\ell^\top \mathbf{A}(\mathbf{x})^\top \mathbf{v}_\ell + \mathbf{b}(\mathbf{x})^\top \mathbf{w} - w_0,$$

where  $w_0$  is the dual variable associated with the linear constraint  $\mathbf{z}^\top \mathbf{u} + u = 1$ . However, the maximization problem above is not convex in  $(\mathbf{z}, \{\mathbf{v}_\ell\}_\ell, \mathbf{w}, w_0)$  due to the product of variable  $z_\ell w_0$  in the constraint  $z_\ell h^*(\frac{\mathbf{v}_\ell}{z_\ell}) \leq z_\ell w_0$  for every  $\ell = 1, \dots, L$ .

Since the safe approximation in Proposition 4 is an adaptive robust linear constraint, we can then derive tractable reformulations by restricting our attention to specific adaptive policies. For instance, if we consider only static policies, we obtain the following safe approximations of the inequalities (23b) and (23a) which are equivalent to Corollary 1:

**Corollary 2** Under Assumption 3, a safe approximation of (1) is:

$$\sum_{\ell \in [L]} z_\ell h_\infty(\mathbf{A}(\mathbf{x})\mathbf{e}_\ell) + h(\mathbf{b}(\mathbf{x})) \leq 0, \quad \forall \mathbf{z} \in \mathcal{Z}.$$

**Proof**

$$\begin{aligned} \exists (\mathbf{u}, u) \in \mathbb{R}_+^{L+1}, \text{ s.t. } \sum_{\ell \in [L]} z_\ell u_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{u_\ell}\right) + u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right) \leq 0, \quad \forall \mathbf{z} \in \mathcal{Z}, \\ \mathbf{z}^\top \mathbf{u} + u = 1, \quad \forall \mathbf{z} \in \mathcal{Z}. \end{aligned}$$

Since  $\mathcal{Z}$  is full dimensional by Assumption 3, the equality constraint implies  $\mathbf{u} = \mathbf{0}$  and  $u = 1$ .  $\square$

**Remark 6** For the safe approximation (14) to be meaningful, we need the recession function of  $h$ ,  $h_\infty(\cdot)$  to be finite. If this is not the case - if  $h$  is quadratic for instance - then, by decomposing  $\mathbf{A}(\mathbf{x})\mathbf{z} + \mathbf{b}(\mathbf{x})$  into  $\sum_{\ell} z_\ell \mathbf{A}(\mathbf{x})\mathbf{e}_\ell + \mathbf{b}(\mathbf{x}) + \mathbf{0}$  and applying the same line of reasoning as for Proposition 4, we can show that

$$\begin{aligned} \exists (\mathbf{u}, u) \in \mathbb{R}_+^{L+1}, \text{ s.t. } \sum_{\ell \in [L]} z_\ell u_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{u_\ell}\right) + u h\left(\frac{\mathbf{b}(\mathbf{x})}{u}\right) \leq 0, \quad \forall \mathbf{z} \in \mathcal{Z}, \\ \mathbf{z}^\top \mathbf{u} + u \leq 1, \quad \forall \mathbf{z} \in \mathcal{Z}, \end{aligned}$$

also constitutes a safe approximation of (1) under the additional assumption that  $h(\mathbf{0}) = 0$ . However, now, the inequality constraint does not force  $\mathbf{u} = \mathbf{0}$  so it does not involve the recession function of  $h$ .

## A.2 Probabilistic guarantees

The main advantage of the safe approximation obtained with Corollary 1-2 is that the resulting constraint is a robust constraint with the same uncertainty set  $\mathcal{Z}$  for which probabilistic guarantees can be derived. In order to provide such guarantees, distributional assumptions on the “true” uncertain parameter are needed. However, Assumption 3 requires the uncertainty set  $\mathcal{Z}$  to lie in the non-negative orthant, an assumption which is often satisfied in practice once the original uncertain parameter is shifted or lifted. Accordingly, we derive probabilistic guarantees for these two cases.

For clarity, we will refer to this original uncertain vector as  $\mathbf{u}$  and identify random variables with tildes ( $\tilde{\cdot}$ ).

### A.2.1 Shifted uncertainty set

We first consider the case where the uncertainty set is shifted and make the following assumption:

**Assumption 4** We assume that

- (i) the uncertainty set  $\mathcal{Z} \subseteq \mathbb{R}_+^L$  can be decomposed into  $\mathcal{Z} = \mathbf{z}_0 + \mathcal{U}$ , where  $\mathcal{U}$  is full dimensional and contains  $\mathbf{0}$  in its interior;
- (ii) the random vector  $\tilde{\mathbf{z}}$  can be decomposed into  $\tilde{\mathbf{z}} = \mathbf{z}_0 + \tilde{\mathbf{u}}$ , where  $\mathbf{z}_0$  is a deterministic vector and  $\tilde{\mathbf{u}}$  is a random vector whose coordinates are  $L$  independent sub-Gaussian random variables with parameter 1 (see Definition 1).

In practice, the uncertain parameter  $\mathbf{z}$  is often decomposed into  $\mathbf{z} = \hat{\mathbf{z}} + \mathbf{u}$ , where  $\hat{\mathbf{z}}$  is the nominal value and  $\mathbf{u}$  the (uncertain) deviation from the nominal. Then, an uncertainty set for  $\mathbf{u}$  is built,  $\mathcal{U}$ , and  $\mathbf{z}$  is taken in the uncertainty set  $\hat{\mathbf{z}} + \mathcal{U}$ . Assumption 4 mirrors this modelling process with the additional caveat that the uncertainty set is further shifted by  $\mathbf{z}_0 - \hat{\mathbf{z}}$  in order to guarantee that  $\mathcal{Z} \subseteq \mathbb{R}_+^L$ .

From a probability distribution standpoint, Assumption 4 requires the coordinates of  $\tilde{\mathbf{u}}$  to be sub-Gaussian, i.e.,

**Definition 1** [17, Definition 1.2 in] A random variable  $\tilde{u} \in \mathbb{R}$  is said to be *sub-Gaussian* with parameter  $\sigma^2$ , denoted  $\tilde{u} \sim \text{subG}(\sigma^2)$ , if  $\tilde{u}$  is centered, i.e.,  $\mathbb{E}[\tilde{u}] = 0$ , and for all  $s \in \mathbb{R}$ ,

$$\mathbb{E} \left[ e^{s\tilde{u}} \right] \leq e^{\frac{s^2\sigma^2}{2}}.$$

Requiring  $\tilde{\mathbf{u}}$  to be sub-Gaussian or, equivalently, to have Gaussian-type tail probabilities, is a fairly general assumption. In particular, standard Gaussian random variables or random variables with mean 0 and bounded between  $-1$  and  $1$  are special cases of sub-Gaussian random variables with parameter 1; see [17, Lemma 1.8]. From a bound on the moment generating function, one can derive large deviation bounds. More precisely, if  $\tilde{u} \sim \text{subG}(\sigma^2)$ , then for any  $t > 0$ ,  $\mathbb{P}(\tilde{u} > t) \leq \exp(-t^2/(2\sigma^2))$  [17, Lemma 1.3].

Under this set of assumptions, we derive the following probabilistic guarantees.



**Proposition 5** Under Assumption 3 and 4, if  $\mathbf{x}$  satisfies

$$\mathbf{z}^\top \mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) + h(\mathbf{b}(\mathbf{x})) \leq 0, \quad \forall \mathbf{z} \in \mathcal{Z},$$

with  $\mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) = (h_\infty(\mathbf{A}(\mathbf{x})\mathbf{e}_1), \dots, h_\infty(\mathbf{A}(\mathbf{x})\mathbf{e}_L))^\top$ , then

$$\begin{aligned} \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0) &\leq \exp\left(-\frac{\delta^*(\mathbf{h}_\infty(\mathbf{A}(\mathbf{x}))|\mathcal{U})^2}{2\|\mathbf{h}_\infty(\mathbf{A}(\mathbf{x}))\|_2^2}\right) + p_0 \\ &\leq \exp\left(-\frac{1}{2}\rho(\mathcal{U})^2\right) + p_0, \end{aligned}$$

where

- a)  $\delta^*(\mathbf{y}|\mathcal{U}) := \max_{\mathbf{u} \in \mathcal{U}} \mathbf{y}^\top \mathbf{u}$  is the support function of the set  $\mathcal{U}$ ,
- b)  $\rho(\mathcal{U}) := \min_{\mathbf{y}: \|\mathbf{y}\|_2=1} \delta^*(\mathbf{y}|\mathcal{U})$  denotes its robust complexity [10],
- c) and  $p_0 \geq \mathbb{P}(\tilde{\mathbf{z}} \not\geq \mathbf{0})$ .

Proposition 5 displays a posteriori and a priori probabilistic guarantees, i.e., a bound that depends on a specific robust solution  $\mathbf{x}$  and one that only depends on the uncertainty set  $\mathcal{U}$ . In both cases, it involves  $p_0 \geq \mathbb{P}(\tilde{\mathbf{z}} \not\geq \mathbf{0}) = \mathbb{P}(\tilde{\mathbf{u}} \not\geq -\mathbf{z}_0)$ , i.e., the probability that a random uncertain parameter  $\tilde{\mathbf{z}}$  does not belong to the non-negative orthant. In cases where  $\tilde{\mathbf{z}}$  is known or assumed to have bounded support—for instance  $\tilde{\mathbf{z}} \in [-1, 1]^L$  almost surely—then  $\mathbf{z}_0$  can be chosen large enough so that  $\mathbb{P}(\tilde{\mathbf{z}} \not\geq \mathbf{0}) = 0$ . Otherwise, for a general  $\mathbf{z}_0 \geq \mathbf{0}$ , a union bound inequality yields

$$\mathbb{P}(\tilde{\mathbf{z}} \not\geq \mathbf{0}) = \mathbb{P}(\tilde{\mathbf{u}} \not\geq -\mathbf{z}_0) \leq \sum_{i=1}^L \mathbb{P}(\tilde{u}_i \not\geq -z_{0i}) \leq \sum_{i=1}^L \exp(-z_{0i}^2/2),$$

where the last inequality follows from the fact that  $\tilde{u}_i \sim \text{subG}(1)$ . Hence,  $p_0$  is a valid upper bound on  $\mathbb{P}(\tilde{\mathbf{z}} \not\geq \mathbf{0})$  and can be as low as 0 and in general depends on the shifting vector  $\mathbf{z}_0$ .

The a priori guarantees in Proposition 5 depend on the uncertainty set through its robust complexity,  $\rho(\mathcal{U})$ , introduced by Bertsimas et al. [10]. We report closed-form expressions of  $\rho(\mathcal{U})$ , or at least a lower bound, for commonly used uncertainty set in Table 3.

**Proof** We consider a robust solution  $\mathbf{x}$ . With support function notations,  $\mathbf{x}$  satisfies

$$\delta^*(\mathbf{h}_\infty(\mathbf{A}(\mathbf{x}))|\mathcal{Z}) + h(\mathbf{b}(\mathbf{x})) \leq 0.$$

We denote  $\mathcal{A}$  the event that  $\tilde{\mathbf{z}} \geq \mathbf{0}$ . Hence,  $\mathbb{P}(\mathcal{A}^c) = \mathbb{P}(\tilde{\mathbf{z}} \not\geq \mathbf{0}) \leq p_0$ , by definition of  $p_0$ . Conditioned on  $\mathcal{A}$ , the assumptions of Lemma 1 are satisfied, leading to

$$h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) \leq \tilde{\mathbf{z}}^\top \mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) + h(\mathbf{b}(\mathbf{x})),$$

**Table 3** Valid lower bound on robust complexity of  $\mathcal{U}$ , defined as  $\rho(\mathcal{U}) := \min_{\mathbf{y}: \|\mathbf{y}\|_2=1} \max_{\mathbf{u} \in \mathcal{Z}} \mathbf{y}^\top \mathbf{u}$ , for some common uncertainty sets. Instances denoted by a \* are valid under the assumption that the true uncertain parameter  $\tilde{\mathbf{u}}$  satisfies  $\|\tilde{\mathbf{u}}\|_\infty \leq 1$ . Results are taken from Table 2 in Bertsimas et al. [10]

Uncertainty set	Definition	Safe approximation of $\rho(\mathcal{U})$
Norm-set	$\{\mathbf{u} : \ \mathbf{u}\ _p \leq \Gamma\}$	$\begin{cases} \Gamma, & \text{if } p \geq 2, \\ \Gamma L^{1/2-1/p}, & \text{if } p \leq 2. \end{cases}$
Budget set*	$\{\mathbf{u} : \ \mathbf{u}\ _\infty \leq 1, \ \mathbf{z}\ _1 \leq \Gamma\}$	$\Gamma/\sqrt{L}$
Box-Ellipsoidal set*	$\{\mathbf{u} : \ \mathbf{u}\ _\infty \leq 1, \ \mathbf{z}\ _2 \leq \Gamma\}$	$\Gamma$
$\ell_\infty + \ell_1$ set	$\{\mathbf{u}_1 + \mathbf{u}_2 : \ \mathbf{u}_1\ _\infty \leq \rho_1, \ \mathbf{u}_2\ _1 \leq \rho_2\}$	$\rho_1 + \rho_2/\sqrt{L}$
$\ell_\infty + \ell_2$ set	$\{\mathbf{u}_1 + \mathbf{u}_2 : \ \mathbf{u}_1\ _\infty \leq \rho_1, \ \mathbf{u}_2\ _2 \leq \rho_2\}$	$\rho_1 + \rho_2$
Polyhedral set	$\{\mathbf{u} : \mathbf{D}\mathbf{u} \leq \mathbf{d}\}$	$\min_i \frac{d_i}{\ \mathbf{D}^\top \mathbf{e}_i\ _2}$

and

$$\begin{aligned}
 \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0 \mid \mathcal{A}) &\leq \mathbb{P}(\tilde{\mathbf{z}}^\top \mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) > \delta^*(\mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) \mid \mathcal{Z})) \\
 &= \mathbb{P}(\tilde{\mathbf{u}}^\top \mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) > \delta^*(\mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) \mid \mathcal{U})) \\
 &\leq \exp\left(-\frac{\delta^*(\mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) \mid \mathcal{U})^2}{2\|\mathbf{h}_\infty(\mathbf{A}(\mathbf{x}))\|_2^2}\right),
 \end{aligned}$$

where the last inequality follows from the fact that, by affine transformation of sub-Gaussian random variables,  $\tilde{\mathbf{u}}^\top \mathbf{h}_\infty(\mathbf{A}(\mathbf{x}))$  is itself sub-Gaussian with parameter  $\|\mathbf{h}_\infty(\mathbf{A}(\mathbf{x}))\|_2^2$ . By definition of the robust complexity, for any vector  $\mathbf{v}$ ,

$$\frac{\delta^*(\mathbf{v} \mid \mathcal{U})}{\|\mathbf{v}\|_2} \geq \min_{\mathbf{v}: \|\mathbf{v}\|_2=1} \delta^*(\mathbf{v} \mid \mathcal{U}) = \rho(\mathcal{U}),$$

so that we obtain

$$\mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0 \mid \mathcal{A}) \leq \exp\left(-\frac{1}{2}\rho(\mathcal{U})^2\right).$$

All things considered,

$$\begin{aligned}
 \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0) &= \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0 \mid \mathcal{A}) \mathbb{P}(\mathcal{A}) \\
 &\quad + \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0 \mid \mathcal{A}^c) \mathbb{P}(\mathcal{A}^c) \\
 &\leq \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0 \mid \mathcal{A}) + \mathbb{P}(\mathcal{A}^c),
 \end{aligned}$$

and the result follows.  $\square$

**Remark 7** As previously suggested, if the recession function of  $h$  is not finite, we can further assume that  $h(\mathbf{0}) = 0$  and show that

$$\begin{aligned} \exists(v, v) \in \mathbb{R}_+^{L+1}, \text{ s.t. } \sum_{\ell \in [L]} z_\ell v_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{v_\ell}\right) + v h\left(\frac{\mathbf{b}(\mathbf{x})}{v}\right) \leq 0, \forall \mathbf{z} \in \mathcal{Z}, \\ \mathbf{z}^\top \mathbf{v} + v \leq 1, \forall \mathbf{z} \in \mathcal{Z}, \end{aligned}$$

constitutes a valid safe approximation of (1). If  $\mathbf{x}$  satisfies such a safe approximation, then a similar line of reasoning yields (proof omitted)

$$\begin{aligned} \mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{z}} + \mathbf{b}(\mathbf{x})) > 0) &\leq \exp\left(-\frac{\delta^*(\mathbf{g}(\mathbf{x}, \mathbf{v})|\mathcal{U})^2}{2\|\mathbf{g}(\mathbf{x}, \mathbf{v})\|_2^2}\right) + \exp\left(-\frac{\delta^*(\mathbf{v}|\mathcal{U})^2}{2\|\mathbf{v}\|_2^2}\right) + p_0 \\ &\leq 2 \exp\left(-\frac{1}{2}\rho(\mathcal{U})^2\right) + p_0, \end{aligned}$$

with  $\mathbf{g}(\mathbf{x}, \mathbf{v})^\top \mathbf{e}_\ell = v_\ell h\left(\frac{\mathbf{A}(\mathbf{x})\mathbf{e}_\ell}{v_\ell}\right)$ .

## A.2.2 Lifted uncertainty set

We now consider the case where  $\mathcal{Z}$  is a lifted version of an original uncertainty set. Accordingly, we assume the following:

**Assumption 5** We assume that

- (i) the uncertainty set  $\mathcal{Z} \subseteq \mathbb{R}_+^{2L}$  is of the form  $\mathcal{Z} = \{(\mathbf{z}^+, \mathbf{z}^-) \in \mathbb{R}_+^L \times \mathbb{R}_+^L \mid \exists \mathbf{u} \in \mathcal{U} : \mathbf{z}^+ = \max(\mathbf{u}, \mathbf{0}), \mathbf{z}^- = \max(-\mathbf{u}, \mathbf{0})\}$ , where  $\mathcal{U}$  is full dimensional and contains  $\mathbf{0}$  in its interior;
- (ii) the random vector  $\tilde{\mathbf{z}}$  can be decomposed into  $\tilde{\mathbf{z}} = (\max(\tilde{\mathbf{u}}, \mathbf{0}), \max(-\tilde{\mathbf{u}}, \mathbf{0})) \in \mathbb{R}_+^{2L}$ , where the coordinates of  $\tilde{\mathbf{u}} \in \mathbb{R}^L$  are  $L$  independent sub-Gaussian random variables with parameter 1.

Under this assumption, we can write Corollary 1–2 as:

$$(\mathbf{z}^+)^\top \mathbf{h}_\infty(\mathbf{A}(\mathbf{x})) + (\mathbf{z}^-)^\top \mathbf{h}_\infty(-\mathbf{A}(\mathbf{x})) + h(\mathbf{b}(\mathbf{x})) \leq 0, \forall (\mathbf{z}^+, \mathbf{z}^-) \in \mathcal{Z}.$$

Since  $\mathbf{z}^+$  and  $\mathbf{z}^-$  cannot be simultaneously positive, the constraint above is equivalent to

$$\sum_{\ell} |u_\ell| \max[h_\infty(\pm \mathbf{A}(\mathbf{x})\mathbf{e}_\ell)] + h(\mathbf{b}(\mathbf{x})) \leq 0, \forall \mathbf{u} \in \mathcal{U}.$$

Despite additional technicalities due to the fact that the coordinates of  $\mathbf{z}^\pm$  are defined as the positive (resp. negative) parts of sub-Gaussian random variables, we obtain similar guarantees.

**Proposition 6** Define  $|\mathcal{U}| := \{|\mathbf{u}| : \mathbf{u} \in \mathcal{U}\}$ ,  $\mathbf{g}(\mathbf{A}, \mathbf{x}) := [\max h_\infty(\pm \mathbf{A}(\mathbf{x})\mathbf{e}_\ell)]_{\ell=1,\dots,L}$  and  $\alpha = \sqrt{L}$ . Consider a vector  $\mathbf{x}$  satisfying

1. If  $t := \frac{\delta^*(\mathbf{g}(\mathbf{A}, \mathbf{x})|\mathcal{U}|)}{\alpha \|\mathbf{g}(\mathbf{A}, \mathbf{x})\|_2} > 1$ , and under Assumption 5, then

$$\mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{u}} + \mathbf{b}(\mathbf{x})) > 0) \leq \sqrt{e}t \exp\left(-\frac{t^2}{2}\right).$$

2. If  $\rho(|\mathcal{U}|) > \alpha$  and under Assumption 5, then

$$\mathbb{P}(h(\mathbf{A}(\mathbf{x})\tilde{\mathbf{u}} + \mathbf{b}(\mathbf{x})) > 0) \leq \frac{\sqrt{e}}{\alpha} \rho(|\mathcal{U}|) \exp\left(-\frac{1}{2\alpha^2} \rho(|\mathcal{U}|)^2\right).$$

Proposition 6 recovers exactly Theorem 2 from Bertsimas et al. [9], under the weaker assumption that  $\mathbf{u}_j$ 's are independent sub-Gaussian random variables (instead of independent Gaussian random variables) and for any uncertainty set  $\mathcal{U}$ . In particular, our proof relies on a novel generalization of the Hanson-Wright inequality to sub-Gaussian random variables, which is of independent interest (see the Online Supplement, Theorem OS.2 and Corollary OS.4). However, the constant  $\alpha$  in Proposition 6 is weaker than the ones obtained by Bertsimas et al. [9]. Indeed, while our bound is generic, they provide  $h$ -specific values for  $\alpha$ , which we could use to tighten our bound as well.

Before proving Proposition 6, we establish a technical lemma.

**Lemma 2** For any  $\mathbf{x}$  and  $\mathbf{z}$ , we have

$$\max(h_\infty(\pm \mathbf{A}(\mathbf{x})\mathbf{z})) \leq \alpha \|\mathbf{H}(\mathbf{A}, \mathbf{x})\mathbf{z}\|_2.$$

for  $\alpha = \sqrt{L}$  and  $\mathbf{H}(\mathbf{A}, \mathbf{x})$  the  $L \times L$  diagonal matrix whose diagonal coefficients are  $\max[h_\infty(\pm \mathbf{A}(\mathbf{x})\mathbf{e}_\ell)]$ .

**Proof** Since  $h_\infty$  is positively homogeneous with  $h_\infty(\mathbf{0}) = 0$ , it is sub-additive and

$$\begin{aligned} h_\infty(\mathbf{A}(\mathbf{x})\mathbf{z}) &\leq \sum_{\ell} h_\infty(z_\ell \mathbf{A}(\mathbf{x})\mathbf{e}_\ell) && \text{(Sub-additivity)} \\ &\leq \sum_{j=1}^L |z_j| h_\infty(\epsilon_\ell \mathbf{A}(\mathbf{x})\mathbf{e}_\ell) && \text{with } \epsilon_\ell = \text{sign}(z_\ell) \\ &\leq \sum_{j=1}^L |z_j| \max[h_\infty(\pm \mathbf{A}(\mathbf{x})\mathbf{e}_\ell)] \\ &\leq \sqrt{L} \sqrt{\sum_{j=1}^L \max[h_\infty(\pm \mathbf{A}(\mathbf{x})\mathbf{e}_\ell)]^2 z_j^2} && \text{(Cauchy-Schwarz).} \end{aligned}$$

□

Finally, let us prove the Proposition 6.

**Proof** 1. Using the safe approximation

$$h(A(x)\tilde{u} + b(x)) \leq h_\infty(A(x)\tilde{u}) + h(b(x)) \leq \max[h_\infty(\pm A(x)\tilde{u})] + h(b(x)),$$

together with

$$\delta^*(g(A, x) | \mathcal{U}|) + h(b(x)) \leq 0,$$

we have

$$\mathbb{P}(h(A(x)\tilde{u} + b(x)) > 0) \leq \mathbb{P}(\max[h_\infty(\pm A(x)\tilde{u})] > \delta^*(g(A, x) | \mathcal{U}|)).$$

Applying Lemma 2, we eventually get

$$\mathbb{P}(h(A(x)\tilde{z} + b(x)) > 0) \leq \mathbb{P}(\alpha \|H(A, x)\tilde{u}\|_2 > \delta^*(g(A, x) | \mathcal{U}|)).$$

Both sides of the inequality being positive, we compare their squares and the result follows from the generalized Hanson-Wright inequality (Corollary OS.4) with  $t = \delta^*(g(A, x) | \mathcal{U}|)^2 / \alpha^2$ .

2. The function  $t \mapsto \sqrt{et} \exp(-t^2/2)$  is monotonically decreasing over  $[1, \infty)$ .

Denoting  $t(x) = \frac{\delta^*(g(A, x) | \mathcal{U}|)}{\alpha \|g(A, x)\|_2}$ , we have  $t(x) \geq \rho(|\mathcal{U}|)/\alpha$ .

□

**Remark 8** A tighter, yet more overwhelming, bound can be obtained by invoking the generalized Hanson-Wright inequality stated in Theorem OS.2 instead of Corollary OS.4.

**Remark 9** We recover exactly Theorem 2 of Bertsimas et al. [9], under the weaker assumption that  $u_j$ 's are independent sub-Gaussian random variables (instead of independent Gaussian random variables) and for any uncertainty set  $\mathcal{U}$ . However, the generic constant  $\alpha$  we obtain is weaker. Indeed, Bertsimas et al. [9] provide an  $h$ -specific proof of Lemma 2 and are able to derive tighter values of  $\alpha$  in these special cases, which could in turn be used to tighten our bound as well.

## B Derivations for the robust control example

Constraint (20) is of the form

$$h(A(x)z + b(x)) \leq t - \|x\|_2^2, \quad \forall z \in \mathcal{Z},$$

with

$$h\left(\begin{bmatrix} u \\ \tilde{u} \end{bmatrix}\right) = u^\top u + \tilde{u}, \quad A(x) = \begin{bmatrix} C^{1/2} \\ 2x^\top F + 2h^\top \end{bmatrix}, \quad b(x) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

In this case, since there are no linear constraints in the definition of  $\text{dom } h^*$ ,  $\Theta_1 = \Theta_2$  in the hierarchy of safe approximations presented in Proposition 2.

### B.1 Ellipsoidal uncertainty set

We first consider the raw description of the uncertainty set  $\mathcal{Z} = \{z : \|z\|_2 \leq \gamma\}$  and consider the bilinear reformulation (2)

$$\sup_{(w, \bar{w}, w_0, z, V, \bar{v}, v_0) \in \Theta} \left[ \text{Tr}(C^{1/2}V) + 2x^\top F\bar{v} + 2h^\top \bar{v} - w_0 \right] \leq t - \|x\|_2^2,$$

with

$$\Theta = \left\{ (w, \bar{w}, w_0, z, V, \bar{v}, v_0) \left| \begin{array}{ll} \|z\|_2 \leq \gamma, & (z \in \mathcal{Z}) \\ \bar{w} = 1, & ((w, \bar{w}) \in \text{dom } h^*) \\ 0 \leq w_0, \frac{1}{4}\|w\|_2^2 \leq w_0, & (h^*(w, \bar{w}) \leq w_0) \\ V = wz^\top, & \\ \bar{v} = \bar{w}z, & \\ v_0 = w_0z & \end{array} \right. \right\}.$$

After careful inspection, we can eliminate the variables  $\bar{w}(= 1)$  and  $\bar{v}(= z)$  and consider instead

$$\sup_{(w, w_0, z, V, v_0) \in \Theta} \left[ \text{Tr}(C^{1/2}V) + 2x^\top Fz + 2h^\top z - w_0 \right] \leq t - \|x\|_2^2,$$

with

$$\Theta = \left\{ (w, w_0, z, V, v_0) \left| \begin{array}{l} \|z\|_2 \leq \gamma, \\ 0 \leq w_0, \frac{1}{4}\|w\|_2^2 \leq w_0, \\ V = wz^\top, \\ v_0 = w_0z \end{array} \right. \right\}.$$

However, in the absence of linear inequalities in the description of  $\mathcal{Z}$  and  $\text{dom } h^*$ , we have  $\Theta_0 = \Theta_1 = \Theta_2$  and the resulting safe approximations are uninformative. Yet, we can apply the third level of approximation,

$$\Theta_3 = \left\{ (w, w_0, z, V, v_0) \left| \begin{array}{l} \|z\|_2 \leq \gamma, \\ 0 \leq w_0, \frac{1}{4}\|w\|_2^2 \leq w_0, \\ \frac{1}{4}\|V\|_2^2 \leq \gamma^2 w_0, \\ \|v_0\|_2 \leq \gamma w_0 \end{array} \right. \right\},$$

and eventually get the following robust counterpart

$$2\gamma\|F^\top x + h\|_2 + \gamma^2\|C\|_2 \leq t - \|x\|_2^2.$$

This is equivalent to upper-bounding the strongly convex part in the original constraint by  $\gamma^2 \|C\|_2$  and take the robust counterpart of the robust *linear* constraint. This is the approach chosen by Bertsimas and Brown [6], although with different derivations.

## B.2 Shifted uncertainty set

We enrich the definition of  $\mathcal{Z}$  with redundant linear constraints, i.e., we consider

$$\mathcal{Z} = \{z : \|z\|_2 \leq \gamma, \underline{z} \leq z \leq \bar{z}\},$$

with  $\bar{z} = \gamma e$  and  $\underline{z} = -\gamma e$ . As a result,  $z - \underline{z} \geq \mathbf{0}$  and we can apply Corollary 1 for the shifted uncertainty  $-\underline{z} + \mathcal{Z} \subseteq \mathbb{R}_+^I$ . Namely, we introduce additional decision variables  $u, u \geq 0$  such that  $u^\top (z - \underline{z}) + u \leq 1, \forall z$ , i.e.,  $\gamma \|u\|_2 - u^\top \underline{z} + u \leq 1$ , and such that

$$-\sum_i \bar{z}_i \frac{C_{ii}}{u_i} + \gamma \left\| 2F^\top x + 2h + \left[ \frac{C_{ii}}{u_i} \right]_i \right\|_2 + \frac{\bar{z}^\top C \bar{z}}{u} \leq t - \|x\|_2.$$

Alternatively,  $\bar{z} - z \geq \mathbf{0}$  and we can apply Corollary 1 for the shifted uncertainty  $\bar{z} - \mathcal{Z} \subseteq \mathbb{R}_+^I$ . Namely, we introduce additional decision variables  $u, u \geq 0$  such that  $u^\top (\bar{z} - z) + u \leq 1, \forall z$ , i.e.,  $u^\top \bar{z} + \gamma \|u\|_2 + u \leq 1$ , and such that

$$\sum_i \bar{z}_i \frac{C_{ii}}{u_i} + \gamma \left\| 2F^\top x + 2h - \left[ \frac{C_{ii}}{u_i} \right]_i \right\|_2 + \frac{\bar{z}^\top C \bar{z}}{u} \leq t - \|x\|_2.$$

Finally, we can also employ the bilinear reformulation (2):

$$\sup_{(w, w_0, z, V, v_0) \in \Theta} \left[ \text{Tr}(C^{1/2} V) + 2x^\top F z + 2h^\top z - w_0 \right] \leq t - \|x\|_2^2,$$

with

$$\Theta = \left\{ (w, w_0, z, V, v_0) \left| \begin{array}{l} \underline{z} \leq z, z \leq \bar{z}, \|z\|_2 \leq \gamma, (z \in \mathcal{Z}) \\ \frac{1}{4} \|w\|_2^2 \leq w_0, (h^*(w, \bar{w}) \leq w_0) \\ V = wz^\top, \\ v_0 = w_0 z \end{array} \right. \right\},$$

and derive the corresponding hierarchy (Proposition 2):

$$\Theta_0 = \left\{ (w, w_0, z, V, v_0) \left| \begin{array}{l} \underline{z} \leq z, z \leq \bar{z}, \|z\|_2 \leq \gamma, (z \in \mathcal{Z}) \\ \frac{1}{4} \|w\|_2^2 \leq w_0, (h^*(w, \bar{w}) \leq w_0) \\ w_0 \underline{z} \leq v_0 \leq w_0 \bar{z}, \|v_0\|_2 \leq \gamma w_0, \end{array} \right. \right\},$$

$$\Theta_1 = \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \left| \begin{array}{l} (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta_0, \\ \frac{1}{4}(\bar{z}_k - z_k) \left\| \frac{\bar{z}_k \mathbf{w} - \mathbf{V} \mathbf{e}_k}{\bar{z}_k - z_k} \right\|_2^2 \leq w_0 \bar{z}_k - v_{0k}, \\ \frac{1}{4}(z_k - \bar{z}_k) \left\| \frac{\mathbf{V} \mathbf{e}_k - \bar{z}_k \mathbf{w}}{z_k - \bar{z}_k} \right\|_2^2 \leq v_{0k} - w_0 \bar{z}_k \end{array} \right. \right\},$$

$$\Theta_2 = \Theta_1,$$

$$\Theta_3 = \left\{ (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \left| \begin{array}{l} (\mathbf{w}, w_0, \mathbf{z}, \mathbf{V}, \mathbf{v}_0) \in \Theta_2, \\ \frac{1}{4} \|\mathbf{V}\|_F^2 \leq \gamma^2 w_0. \end{array} \right. \right\}.$$

### B.3 Lifted uncertainty set

Another way to introduce meaningful constraints is to lift the uncertainty set into positive and negative part, i.e., consider  $\{(\mathbf{z}^+, \mathbf{z}^-) \in \mathbb{R}_+^{2L} \mid \|\mathbf{z}^+ + \mathbf{z}^-\|_2 \leq \gamma\}$ . In this case, we can apply Corollary 1 for the lifted uncertainty set and introduce additional decision variables  $\mathbf{u}^+, \mathbf{u}^-, u^+, u^- \geq 0$  such that

$$\gamma \|\max(\mathbf{u}^+, \mathbf{u}^-)\|_2 + u^+ + u^- \leq 1,$$

$$\gamma \left\| \max \left( \left[ \frac{C_{ii}}{u_{+i}} \right]_i + 2\mathbf{F}^\top \mathbf{x} + 2\mathbf{h}, \left[ \frac{C_{ii}}{u_{-i}} \right]_i - 2\mathbf{F}^\top \mathbf{x} - 2\mathbf{h} \right) \right\|_2 \leq t - \|\mathbf{x}\|_2^2.$$

We can also apply the bilinear formulation and enforce

$$\sup_{(\mathbf{w}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0) \in \Theta} \left[ \text{Tr}(\mathbf{C}^{1/2} \mathbf{V}^+) - \text{Tr}(\mathbf{C}^{1/2} \mathbf{V}^-) + 2(\mathbf{F}^\top \mathbf{x} + \mathbf{h})^\top (\mathbf{z}^+ - \mathbf{z}^-) - w_0 \right]$$

to be less than  $t - \|\mathbf{x}\|_2^2$ , with

$$\Theta = \left\{ (\mathbf{w}, \bar{\mathbf{w}}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0^\pm) \left| \begin{array}{l} \mathbf{0} \leq \mathbf{z}^\pm, \|\mathbf{z}^+ + \mathbf{z}^-\|_2 \leq \gamma, ((\mathbf{z}^+, \mathbf{z}^-) \in \mathcal{Z}) \\ 0 \leq w_0, \frac{1}{4} \|\mathbf{w}\|_2^2 \leq w_0, (h^*(\mathbf{w}, \bar{\mathbf{w}}) \leq w_0) \\ \mathbf{V}^\pm = \mathbf{w} \mathbf{z}^{\pm\top}, \\ \mathbf{v}_0^\pm = w_0 \mathbf{z}^\pm \end{array} \right. \right\},$$

and the corresponding hierarchy is

$$\Theta_0 = \left\{ (\mathbf{w}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0^\pm) \left| \begin{array}{l} \mathbf{0} \leq \mathbf{z}^\pm, \|\mathbf{z}^+ + \mathbf{z}^-\|_2 \leq \gamma, ((\mathbf{z}^+, \mathbf{z}^-) \in \mathcal{Z}) \\ 0 \leq w_0, \frac{1}{4} \|\mathbf{w}\|_2^2 \leq w_0, (h^*(\mathbf{w}, \bar{\mathbf{w}}) \leq w_0) \\ \mathbf{v}_0^\pm \geq \mathbf{0}, \|\mathbf{v}_0^+ + \mathbf{v}_0^-\|_2 \leq \gamma w_0 \end{array} \right. \right\},$$

$$\Theta_1 = \left\{ (\mathbf{w}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0^\pm) \left| \begin{array}{l} (\mathbf{w}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0^\pm) \in \Theta_0 \\ \frac{z_k^\pm}{4} \left\| \frac{1}{z_k^\pm} \mathbf{V}^\pm \mathbf{e}_k \right\|_2^2 \leq v_{0,k}^\pm \end{array} \right. \right\},$$

$$\Theta_2 = \Theta_1,$$



$$\Theta_3 = \left\{ (\mathbf{w}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0^\pm) \mid (\mathbf{w}, w_0, \mathbf{z}^\pm, \mathbf{V}^\pm, \mathbf{v}_0^\pm) \in \Theta_2 \mid \frac{1}{4} \|\mathbf{V}^+ + \mathbf{V}^-\|_F^2 \leq \gamma^2 w_0 \right\}.$$

## C Derivations for the robust geometric optimization example

### C.1 Original formulation

The safe approximation at Level 1 with  $\mathcal{Z}_{box} = \{\mathbf{z} \mid \|\mathbf{z}\|_\infty \leq 1\}$  (instead of  $\mathbf{z}$ ) is:

$$\Theta_{1'} = \left\{ (\mathbf{w}, \mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0) \mid \begin{array}{l} \|\mathbf{z}\|_\infty \leq 1 \\ w_1, w_2 \geq 0, w_1 + w_2 = 1 \\ w_1 \log w_1 + w_2 \log w_2 \leq w_0 \\ \|\mathbf{v}_1\|_\infty \leq w_1, \|\mathbf{v}_2\|_\infty \leq w_2 \\ \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{z} \\ (w_1 - v_{1\ell}) \log \frac{w_1 - v_{1\ell}}{1 - z_\ell} + (w_2 - v_{2\ell}) \log \frac{w_2 - v_{2\ell}}{1 - z_\ell} \\ \leq w_0 - v_{0\ell}, \ell \in [L] \\ (w_1 + v_{1\ell}) \log \frac{w_1 + v_{1\ell}}{1 + z_\ell} + (w_2 + v_{2\ell}) \log \frac{w_2 + v_{2\ell}}{1 + z_\ell} \\ \leq w_0 + v_{0\ell}, \ell \in [L] \end{array} \right\}.$$

The safe approximation at Level 1 is:

$$\Theta_1 = \{(\mathbf{w}, \mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0) \in \Theta_{1'} \mid \|\mathbf{z}\|_2 \leq \gamma\}.$$

The safe approximation at Level 2 is:

$$\Theta_2 = \{(\mathbf{w}, \mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0) \in \Theta_1 \mid \|\mathbf{v}_1\|_2 \leq w_1 \gamma, \|\mathbf{v}_2\|_2 \leq w_2 \gamma\}.$$

The safe approximation at Level 3 is:

$$\Theta_{LMI,2} = \left\{ (\mathbf{w}, \mathbf{z}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_0) \in \Theta_2 \mid \exists \mathbf{z}, \mathbf{W} : \begin{pmatrix} \mathbf{W} & \mathbf{V} & \mathbf{w} \\ \mathbf{V}^\top & \mathbf{z} & \mathbf{z} \\ \mathbf{w}^\top & \mathbf{z}^\top & 1 \end{pmatrix} \succeq 0 \right\},$$

where  $\mathbf{W} \in \mathbb{R}^{3 \times 3}$ ,  $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2]^\top \in \mathbb{R}^{3 \times L}$ ,  $\mathbf{z} \in \mathbb{R}^{L \times L}$  and  $\mathbf{w} = [w_0 \ w_1 \ w_2]^\top$ .

## C.2 Lifted formulation

The safe approximation of the extended approach of Bertsimas and Sim [8] is:

$$\Theta_{\text{BS}} = \left\{ \begin{pmatrix} \mathbf{w}, \mathbf{z}^+, \mathbf{z}^-, \mathbf{v}_1^+, \mathbf{v}_1^- \\ \mathbf{v}_2^+, \mathbf{v}_2^-, \mathbf{v}_0^+, \mathbf{v}_0^- \end{pmatrix} \left| \begin{array}{l} \|\mathbf{z}^+ + \mathbf{z}^-\|_\infty \leq 1, \|\mathbf{z}^+ + \mathbf{z}^-\|_2 \leq \gamma \\ w_1, w_2 \geq 0, w_1 + w_2 = 1 \\ w_1 \log w_1 + w_2 \log w_2 \leq w_0 \\ \|\mathbf{v}_1^+ + \mathbf{v}_1^-\|_\infty \leq w_1, \|\mathbf{v}_2^+ + \mathbf{v}_2^-\|_\infty \leq w_2 \\ \mathbf{v}_1^+ + \mathbf{v}_2^+ = \mathbf{z}^+, \mathbf{v}_1^- + \mathbf{v}_2^- = \mathbf{z}^- \\ v_{1\ell}^+ \log(v_{1\ell}^+/z_\ell^+) + v_{2\ell}^+ \log(v_{2\ell}^+/z_\ell^+) \leq v_{0\ell}^+ \ell \in [L] \\ v_{1\ell}^- \log(v_{1\ell}^-/z_\ell^-) + v_{2\ell}^- \log(v_{2\ell}^-/z_\ell^-) \leq v_{0\ell}^- \ell \in [L] \\ \mathbf{z}^+, \mathbf{z}^-, \mathbf{v}_1^+, \mathbf{v}_1^-, \mathbf{v}_2^+, \mathbf{v}_2^- \in \mathbb{R}_+^L \end{array} \right. \right\}.$$

The safe approximation at Level 1 with  $\mathcal{Z}_{\text{box}}^+ = \{(\mathbf{z}^+, \mathbf{z}^-) \in \mathbb{R}_+^L \times \mathbb{R}_+^L \mid \|\mathbf{z}^+ + \mathbf{z}^-\|_\infty \leq 1\}$  is:

$$\Theta_{1'}^+ = \left\{ \begin{pmatrix} \mathbf{w}, \mathbf{z}^+, \mathbf{z}^-, \mathbf{v}_1^+, \mathbf{v}_1^- \\ \mathbf{v}_2^+, \mathbf{v}_2^-, \mathbf{v}_0^+, \mathbf{v}_0^- \end{pmatrix} \left| \begin{array}{l} \|\mathbf{z}^+ + \mathbf{z}^-\|_\infty \leq 1 \\ w_1, w_2 \geq 0, w_1 + w_2 = 1 \\ w_1 \log w_1 + w_2 \log w_2 \leq w_0 \\ \|\mathbf{v}_1^+ + \mathbf{v}_1^-\|_\infty \leq w_1, \|\mathbf{v}_2^+ + \mathbf{v}_2^-\|_\infty \leq w_2 \\ \mathbf{v}_1^+ + \mathbf{v}_2^+ = \mathbf{z}^+, \mathbf{v}_1^- + \mathbf{v}_2^- = \mathbf{z}^- \\ v_{1\ell}^+ \log(v_{1\ell}^+/z_\ell^+) + v_{2\ell}^+ \log(v_{2\ell}^+/z_\ell^+) \leq v_{0\ell}^+ \ell \in [L] \\ v_{1\ell}^- \log(v_{1\ell}^-/z_\ell^-) + v_{2\ell}^- \log(v_{2\ell}^-/z_\ell^-) \leq v_{0\ell}^- \ell \in [L] \\ (w_1 - v_{1\ell}^+ - v_{1\ell}^-) \log \frac{w_1 - v_{1\ell}^+ - v_{1\ell}^-}{1 - z_\ell^+ - z_\ell^-} + \\ (w_2 - v_{2\ell}^+ - v_{2\ell}^-) \log \frac{w_2 - v_{2\ell}^+ - v_{2\ell}^-}{1 - z_\ell^+ - z_\ell^-} \dots \\ \leq w_0 - v_{0\ell}^+ - v_{0\ell}^-, \ell \in [L] \\ \mathbf{z}^+, \mathbf{z}^-, \mathbf{v}_1^+, \mathbf{v}_1^-, \mathbf{v}_2^+, \mathbf{v}_2^- \in \mathbb{R}_+^L \end{array} \right. \right\}.$$

The safe approximation at Level 1 is:

$$\Theta_1^+ = \left\{ \begin{pmatrix} \mathbf{w}, \mathbf{z}^+, \mathbf{z}^-, \mathbf{v}_1^+, \mathbf{v}_1^- \\ \mathbf{v}_2^+, \mathbf{v}_2^-, \mathbf{v}_0^+, \mathbf{v}_0^- \end{pmatrix} \in \Theta_{1'}^+ \mid \|\mathbf{z}^+ + \mathbf{z}^-\|_2 \leq \gamma \right\}.$$

The safe approximation at Level 2 is:

$$\Theta_2^+ = \left\{ \begin{pmatrix} \mathbf{w}, \mathbf{z}^+, \mathbf{z}^-, \mathbf{v}_1^+, \mathbf{v}_1^- \\ \mathbf{v}_2^+, \mathbf{v}_2^-, \mathbf{v}_0^+, \mathbf{v}_0^- \end{pmatrix} \in \Theta_1^+ \mid \|\mathbf{v}_1^+ + \mathbf{v}_1^-\|_2 \leq w_1\gamma, \|\mathbf{v}_2^+ + \mathbf{v}_2^-\|_2 \leq w_2\gamma \right\}.$$

## References

1. Ben-Ameur, W., Ouorou, A., Wang, G., Zotkiewicz, M.: Multipolar robust optimization. *EURO Journal on Computational Optimization* **6**(4), 395–434 (2018)
2. Ben-Tal, A., Nemirovski, A., Roos, K.: Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM J. Optim.* **13**(2), 535–560 (2002)
3. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust optimization. Princeton University Press, New Jersey, US (2009)

4. Ben-Tal, A., den Hertog, D., Vial, J.P.: Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Program.* **149**(1–2), 265–299 (2015)
5. Bertsimas, D., Biddkhor, H.: On the performance of affine policies for two-stage adaptive optimization: a geometric perspective. *Math. Program.* **153**(2), 577–594 (2015)
6. Bertsimas, D., Brown, D.B.: Constrained stochastic lqc: a tractable approach. *IEEE Trans. Autom. Control* **52**(10), 1826–1841 (2007)
7. Bertsimas, D., Goyal, V.: On the power and limitations of affine policies in two-stage adaptive optimization. *Math. Program.* **134**, 491–531 (2012)
8. Bertsimas, D., Sim, M.: Tractable approximations to robust conic optimization problems. *Math. Program.* **107**(1–2), 5–36 (2006)
9. Bertsimas, D., Pachamanova, D., Sim, M.: Robust linear optimization under general norms. *Oper. Res. Lett.* **32**(6), 510–516 (2004)
10. Bertsimas, D., den Hertog, D., Pauphilet, J.: Probabilistic guarantees in robust optimization. *SIAM J. Optim.* **31**(4), 2893–2920 (2021)
11. Chen, X., Zhang, Y.: Uncertain linear programs: Extended affinely adjustable robust counterparts. *Oper. Res.* **57**(6), 1469–1482 (2009)
12. El Housni, O., Goyal, V.: Beyond worst-case: A probabilistic analysis of affine policies in dynamic optimization. *Adv. Neural Inf. Process. Syst.* **30**, 4756–4764 (2017)
13. El Housni, O., Goyal, V.: On the optimality of affine policies for budgeted uncertainty sets. *Math. Oper. Res.* **46**(2), 674–711 (2021)
14. Hadjiyiannis, M.J., Goulart, P.J., Kuhn, D.: A scenario approach for estimating the suboptimality of linear decision rules in two-stage robust optimization. In: 2011 50th IEEE Conference on Decision and Control and European Control Conference, IEEE, pp 7386–7391 (2011)
15. Hsiung, K.L., Kim, S.J., Boyd, S.: Tractable approximate robust geometric programming. *Optim. Eng.* **9**(2), 95–118 (2008)
16. MOSEK (2020) MOSEK Modeling Cookbook. MOSEK ApS, <https://docs.mosek.com/modeling-cookbook/index.html>
17. Rigollet, P., Hütter, J.C.: High dimensional statistics. Lecture notes for course 18S997 **813**(814):46 (2015), <https://klein.mit.edu/~rigollet/PDFs/RigNotes17.pdf>
18. Roos, E., den Hertog, D., Ben-Tal, A., de Ruiter, F.J.C.T., Zhen, J.: Tractable approximation of hard uncertain optimization problems. Available on Optimization Online (2018)
19. de Ruiter, F.J.C.T., Zhen, J., den Hertog, D.: Dual approach for two-stage robust nonlinear optimization. *Operations Research* (2022). <https://doi.org/10.1287/opre.2022.2289>
20. Zhen, J., De Moor, D., den Hertog, D.: Reformulation-perspectification-technique: an extension of the reformulation-linearization-technique. Available on Optimization Online (2021)
21. Zhen, J., de Ruiter, F.J.C.T., Roos, E., den Hertog, D.: Robust optimization for models with uncertain second-order cone and semidefinite programming constraints. *INFORMS J. Comput.* **34**(1), 196–210 (2022)