Published copy can be found at: https://doi.org/10.1137/21M1390967

Noname manuscript No.
(will be inserted by the editor)

Probabilistic guarantees in Robust Optimization

A general methodology

Dimitris Bertsimas · Dick den Hertog · Jean Pauphilet

Received: date / Accepted: date

Abstract We develop a general methodology to derive probabilistic guarantees for solutions of robust optimization problems. Our analysis applies broadly to any convex compact uncertainty set and to any constraint affected by uncertainty in a concave manner. In particular, we introduce the notion of robust complexity of an uncertainty set, which is a robust analog of the Rademacher or Gaussian complexity encountered in high-dimensional statistics, and which drives our results. For a variety of uncertainty sets of practical interest, we are able to compute it in closed form or derive valid approximations. To the best of our knowledge, our methodology recovers and extends all the results available in the literature. We also derive improved a posteriori bounds, i.e., bounds which depend on the resulting robust solution. Besides being significantly tighter than a priori bounds, a posteriori bounds can notably be computed for any general convex uncertainty set.

Keywords Robust optimization · Support function · Uncertainty set · Concentration inequality

Mathematics Subject Classification (2010) 90C25 · 65K05

1 Introduction

Over the past decades, Robust Optimization (RO) has emerged as a tractable framework to account for uncertainty in optimization problems [see 8, 22, for a review]. Consider a single linear constraint of the form

$$\boldsymbol{a}^\top \boldsymbol{x} \leq b, \tag{1}$$

Dimitris Bertsimas
Operations Research Center, Massachusetts Institute of Technology
77 Massachusetts Avenue, E40-111, Cambridge, MA 02139, USA
E-mail: dbertsim@mit.edu
ORCID:0000-0002-1985-1003

Dick den Hertog
Department of Econometrics and Operations Research, Tilburg University
E-mail: d.denhertog@uvt.nl

Jean Pauphilet
Operations Research Center, Massachusetts Institute of Technology
E-mail: jpauph@mit.edu
ORCID:0000-0001-6352-0984

where the parameter $\boldsymbol{a}$ is subject to uncertainty and $b \in \mathbb{R}$ is certain, without loss of generality. In a robust approach, $\boldsymbol{a}$ is described as a deterministic yet unknown vector belonging to a so-called uncertainty set $\mathcal{U}$ and constraint (1) is imposed to hold for all values of $\boldsymbol{a} \in \mathcal{U}$, i.e., we consider its robust counterpart

$$\boldsymbol{a}^\top \boldsymbol{x} \leq b, \ \forall \boldsymbol{a} \in \mathcal{U}. \tag{2}$$

On the other hand, the stochastic approach makes distributional assumptions on the random variable $\tilde{\boldsymbol{a}}$ and replaces (1) by some stochastic version, such as

$$\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} \leq b\right) \geq 1 - \varepsilon. \tag{3}$$

While the robust constraint (2) is often more tractable from a computational perspective, the stochastic constraint (3) has undeniable theoretical appeal: it provides explicit control on the risk of constraint violation ε. In an attempt to bridge the gap between tractability and theoretical guarantees, a relevant question in the RO literature has been to derive similar probabilistic guarantees for robust constraints. Namely, given $\boldsymbol{x}$ satisfying (2), provide an upper bound on $\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} > b\right)$. Clearly, $\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} > b\right) \leq \mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}} \notin \mathcal{U}\right)$, yet much tighter bounds can be obtained. Ben-Tal and Nemirovski [2] proved that

$$\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} > b\right) \leq e^{-\rho^2/2}$$

for their proposed box-ellipsoidal uncertainty set $\mathcal{U} = \{\boldsymbol{a} : \ \|\boldsymbol{a}\|_\infty \leq 1, \ \|\boldsymbol{a}\|_2 \leq \rho\}$, under the assumption that the coordinates of $\tilde{\boldsymbol{a}}$ are independent bounded random variables with mean 0. The bound above is notoriously independent of the number of uncertain parameters L, which is typically not the case for $\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}} \notin \mathcal{U}\right)$. The budget uncertainty set $\mathcal{U} = \{\boldsymbol{a} : \ \|\boldsymbol{a}\|_\infty \leq 1, \ \|\boldsymbol{a}\|_1 \leq \rho\}$ Bertsimas and Sim [5], on the other hand, provides a dimension-dependent probabilistic guarantee

$$\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} > b\right) \leq e^{-\rho^2/2L},$$

under the same distributional assumptions. Since these two seminal works in the field, results for other classes of uncertainty sets $\mathcal{U}$ or more general constraint, $f(\boldsymbol{a}, \boldsymbol{x}) \leq 0$, have been derived, yet in a disparate and unsystematic way. Attempts have been made to report existing bounds in a unified manner [24, 25, 26, 30] for various uncertainty set. In particular, they provide a catalog of the known a priori and a posteriori bounds for specific uncertainty sets and linear constraints. Little if no work, however, provides a disciplined methodology that not only encompasses existing results as special cases but enables to derive probabilistic guarantees for any uncertainty set $\mathcal{U}$ and any constraint function f. In addition, a clear relationship between the geometry of the uncertainty set and the resulting uncertainty set is still needed to understand the connection between robust optimization and its out-of-sample performance on random instances. Such is the aspiration of the present paper.

1.1 Literature review

Safe approximation of ambiguous chance constraint: To the best of our knowledge, works on safe approximations of scalar chance constraints are the closest to achieve this methodological goal. In this setting, the goal is to reformulate a given chance constraint

$$\mathbb{P}_{\tilde{\boldsymbol{a}}}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} \leq b\right) \leq \varepsilon,$$

with $\varepsilon > 0$. In most cases, the distribution of $\tilde{\boldsymbol{a}}$ is not known precisely but rather assumed to belong to a certain class, and the objective is to reformulate the ambiguous chance constraint

$$\sup_{\mathbb{P}_{\tilde{\boldsymbol{a}}} \in \mathcal{P}} \mathbb{P}_{\tilde{\boldsymbol{a}}} \left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} \leq b \right) \leq \varepsilon, \tag{4}$$

where $\mathcal{P}$ is a class of allowable probability distributions for $\tilde{\boldsymbol{a}}$. Given some assumption on $\mathcal{P}$, the ambiguous chance constraint (4) can be proven equivalent to a robust constraint of the form

$$\boldsymbol{a}^\top \boldsymbol{x} \leq b, \ \forall \boldsymbol{a} \in \mathcal{U}_\varepsilon,$$

where the uncertainty set $\mathcal{U}_\varepsilon$ depends on $\mathcal{P}$. Exact reformulations have notably been derived in cases where $\tilde{\boldsymbol{a}}$ is normally distributed [18, 41, 42] or has known mean and bounded second-order moments [41, 42]. Ben-Tal et al. [3, chapter 2] derive such reformulations for a general class of probability distributions and in a disciplined manner. For example, if the coordinates of $\tilde{\boldsymbol{a}}$ are independent random variables in $[-1, 1]$ and unimodal with respect to 0, meaning that 0 is the only mode of the distribution, then constraint (2) with

$$\mathcal{U}_\varepsilon = \left\{ \boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2 \ : \ \|\boldsymbol{a}_1\|_\infty \leq 1/2, \ \|\boldsymbol{a}_2\|_2 \leq \sqrt{\ln(1/\varepsilon)/6} \right\}$$

yields (3) [3, Theorem 2.4.4 applied to Example 2.4.7.]. This approach, which takes assumptions on the distribution of $\tilde{\boldsymbol{a}}$ as the primitive and provides a corresponding uncertainty set, has two main shortcomings. First, it cannot provide any probabilistic guarantee for the uncertainty set $\mathcal{U}_\varepsilon$ whenever $\tilde{\boldsymbol{a}}$ satisfies different assumptions. Instead, different assumptions on $\tilde{\boldsymbol{a}}$ would lead to a different uncertainty set. In this setting, probabilistic assumptions dictate the type of uncertainty set to use, rather than tractability considerations, and impede modeling. Second, this approach only justifies the use of uncertainty sets involving the Euclidean norm, which is closely related with the notion of variance, but excludes more general uncertainty sets, such as polyhedral or budget uncertainty sets, despite their tractability and wide use in practice. More recently, Bertsimas et al. [9] propose a data-driven extension of this procedure. Starting from admittedly loose assumptions on the distribution of $\tilde{\boldsymbol{a}}$ and given past observations $\boldsymbol{a}^{(1)}, \dots \boldsymbol{a}^{(N)}$, they leverage hypothesis testing to construct an uncertainty set with a probabilistic guarantee of level ε. Their setting is fully data-driven and applies to a general concave robust constraint. However, the shape of the uncertainty set is still dictated by the distributional assumptions made - and the corresponding hypothesis test. Also, their uncertainty sets are calibrated using a single constraint. If multiple constraints are affected by uncertainty, their method fails to provide guarantees for the ones that were not used during calibration.

Distributionally robust optimization: Probabilistic guarantees and concentration results have been the fundamental building blocks in the distributionally robust optimization (DRO) literature. In the DRO setting, reformulations of ambiguous chance constraints (4) have been derived in cases where the ambiguity set $\mathcal{P}$ is constructed from bounds on moments of the distribution [19, 23, 42, 45], or is defined as a ball around the empirical distribution according to phi-divergence [43], the Wasserstein distance [16, 21, 29, 39], or the relative entropy [37]. We refer to Hanasusanto et al. [27] and references therein for a comprehensive review. Despite powerful out-of-sample performance guarantees and moderate conservatism, DRO approaches do not display the same favorable tractability properties as simple robust approaches, especially in the data-driven settings. For instance, computing the Wasserstein distance between two distributions is #P-hard in general [29, Theorem 3].

Scenario approach: A general way to deal with chance constraints is the scenario approach, i.e., generate N samples $\boldsymbol{a}^{(i)}$, $i = 1, \dots, N$ of $\tilde{\boldsymbol{a}}$ and replace the chance constraint by N deterministic constraints $f(\boldsymbol{a}^{(i)}, \boldsymbol{x}) \leq 0$. Assuming that the function f is convex in $\boldsymbol{x}$, Calafiore and Campi [12, 13] prove that this approach yields a feasible solution to (3) with probability $1-\delta$ if

$$N \geq 2L + \frac{2L}{\varepsilon}\log(\tfrac{2}{\varepsilon}) + \frac{2}{\varepsilon}\log(\tfrac{1}{\delta}).$$

This approach has been later refined [14, 31, 33] and extended to ambiguous chance constraints [17, 20]. The scenario approach, yet, suffers from two limitations. First, the number of constraints to be sampled, N, grows linearly in the dimension of the uncertainty, L, which makes the approach difficult to apply to medium-size problems. Second, it assumes that one can sample from the "true" distribution of $\tilde{\boldsymbol{a}}$, which might not be the case in practice.

1.2 Contributions and structure

In this paper, we start from a general robust constraint

$$f(\boldsymbol{a}, \boldsymbol{x}) \leq 0, \ \forall \boldsymbol{a} \in \mathcal{U},$$

which is equivalent to

$$\max_{\boldsymbol{a} \in \mathcal{U}} f(\boldsymbol{a}, \boldsymbol{x}) \leq 0,$$

and derive valid probabilistic guarantees for any solution $\boldsymbol{x}$ satisfying the robust constraint. We will assume that $\mathcal{U}$ is convex and $f(\boldsymbol{a}, \boldsymbol{x})$ is concave in $\boldsymbol{a}$ for any $\boldsymbol{x}$. Under this assumption, the maximization problem above is well-defined and the robust constraint can be reformulated in a tractable way [4]. These tractable reformulations, which we recall in Section 2, have been instrumental in the adoption of robust optimization in practice. They are also crucial to derive bounds on the probability of constraint violation, as we demonstrate in this paper. When $f(\boldsymbol{a}, \boldsymbol{x})$ is convex in $\boldsymbol{a}$ however, an equivalent tractable reformulation of the robust constraint is out of reach. Safe approximations based on scenario sampling [13] or linear approximation [6] have been proposed, but remain significantly less tractable. Correspondingly, the probabilistic guarantees obtained in the convex case require more stringent assumptions on the underlying distribution; see Conjecture 10.1 in [3] for instance. We restrict our attention to uncertainty sets $\mathcal{U}$ of the form

$$\mathcal{U} = \left\{ \boldsymbol{a} \ : \ \exists \boldsymbol{z} \in \mathcal{Z} \text{ s.t. } \boldsymbol{a} = \bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z} \right\},$$

where $\bar{\boldsymbol{a}}$ is the nominal value of $\boldsymbol{a}$, $\mathcal{Z}$ is a given nonempty, fully-dimensional convex and compact set, with $\boldsymbol{0} \in \text{ri}(\mathcal{Z})$, as in Ben-Tal et al. [4], which are not overly restrictive assumptions given the general form of the constraint. In particular, the fully dimensional assumption is without loss of generality given an appropriate matrix $\boldsymbol{P}$. Our main contribution is a general and simple methodology for deriving probabilistic guarantees for solutions of robust optimization problems, which in turn leads to the following results, summarized in Table 1:

- For constraints of the form $f(\boldsymbol{a}(\boldsymbol{z}), \boldsymbol{x}) \leq 0, \ \forall \boldsymbol{z} \in \mathcal{Z}$, where f is concave in $\boldsymbol{a}$, we show that an uncertainty set $\mathcal{Z}$ yields a probabilistic guarantee of

$$\mathbb{P}\left(f(\tilde{\boldsymbol{a}}, \boldsymbol{x}) > 0\right) \leq \exp\left(-\tfrac{1}{2}\rho(\mathcal{Z})^2\right),$$

  where we define the robust complexity of $\mathcal{Z}$, denoted $\rho(\mathcal{Z})$, as

$$\rho(\mathcal{Z}) := \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \max_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{y}^\top \boldsymbol{z}.$$

Table 1: Summary of our main findings. Here, $\delta^\star(\cdot|\mathcal{Z})$ and $\rho(\mathcal{Z})$ denote respectively the support function and the robust complexity of the uncertainty set $\mathcal{Z}$, defined as $\delta^\star(\boldsymbol{y}|\mathcal{Z}) = \max_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{z}^\top \boldsymbol{y}$ and $\rho(\mathcal{Z}) = \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \delta^\star(\boldsymbol{y}|\mathcal{Z})$. For the concave case, $f_\star(\boldsymbol{v},\boldsymbol{x}) := \inf_{\boldsymbol{a}} \boldsymbol{a}^\top \boldsymbol{v} - f(\boldsymbol{a},\boldsymbol{x})$ is often referred to as the concave conjugate of $f(\cdot,\boldsymbol{x})$.

Constraint type	Linear	Concave in $\boldsymbol{a}$
Nominal constraint	$\boldsymbol{a}^\top \boldsymbol{x} \le b$	$f(\boldsymbol{a},\boldsymbol{x}) \le 0$
Robust counterpart	$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x}\|\mathcal{Z}) \le b$	$\exists \boldsymbol{v},\ \bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{v}\|\mathcal{Z}) - f_\star(\boldsymbol{v},\boldsymbol{x}) \le 0$
A priori probabilistic guarantee (Corollary 1, 2)	$\exp\left(-\frac{1}{2}\rho(\mathcal{Z})^2\right)$	$\exp\left(-\frac{1}{2}\rho(\mathcal{Z})^2\right)$
A posteriori probabilistic guarantee (Theorem 1, 2)	$\exp\left(-\frac{(b-\bar{\boldsymbol{a}}^\top \boldsymbol{x})^2}{2\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2}\right)$	$\exp\left(-\frac{(f_\star(\boldsymbol{v},\boldsymbol{x})-\bar{\boldsymbol{a}}^\top \boldsymbol{v})^2}{2\|\boldsymbol{P}^\top \boldsymbol{v}\|_2^2}\right)$

Table 2: Valid lower bound on robust complexity of $\mathcal{Z}$, defined as $\rho(\mathcal{Z}) := \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \max_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{y}^\top \boldsymbol{z}$, for some common uncertainty sets. Instances denoted by a * are valid under the assumption that the true uncertain parameter $\tilde{\boldsymbol{z}}$ satisfies $\|\tilde{\boldsymbol{z}}\|_\infty \le 1$.

Uncertainty set	Definition	Safe approximation of $\rho(\mathcal{Z})$
Norm-set	$\{\boldsymbol{z} : \|\boldsymbol{z}\|_p \le \Gamma\}$	$\begin{cases} \Gamma, & \text{if } p \ge 2, \\ \Gamma L^{1/2-1/p}, & \text{if } p \le 2. \end{cases}$
Budget set*	$\{\boldsymbol{z} : \|\boldsymbol{z}\|_\infty \le 1, \|\boldsymbol{z}\|_1 \le \Gamma\}$	$\Gamma/\sqrt{L}$
Box-Ellipsoidal set*	$\{\boldsymbol{z} : \|\boldsymbol{z}\|_\infty \le 1, \|\boldsymbol{z}\|_2 \le \Gamma\}$	Γ
ℓ_∞+ℓ_1 set	$\{\boldsymbol{z} = \boldsymbol{z}_1 + \boldsymbol{z}_2 : \|\boldsymbol{z}_1\|_\infty \le \rho_1, \|\boldsymbol{z}_2\|_1 \le \rho_2\}$	$\rho_1 + \rho_2/\sqrt{L}$
ℓ_∞+ℓ_2 set	$\{\boldsymbol{z} = \boldsymbol{z}_1 + \boldsymbol{z}_2 : \|\boldsymbol{z}_1\|_\infty \le \rho_1, \|\boldsymbol{z}_2\|_2 \le \rho_2\}$	$\rho_1 + \rho_2$
Polyhedral set	$\{\boldsymbol{z} : \boldsymbol{D}\boldsymbol{z} \le \boldsymbol{d}\}$	$\min_i \frac{d_i}{\|\boldsymbol{D}^\top \boldsymbol{e}_i\|_2}$

This simple statement recovers all a priori bounds available in the literature for simple uncertainty sets and linear constraints, and extends them to the more general case where $f(\boldsymbol{a},\boldsymbol{x})$ is any function concave in $\boldsymbol{a}$. In particular, we recover known probabilistic guarantees for the box-ellipsoidal set of Ben-Tal and Nemirovski [2] and the budget uncertainty set of Bertsimas and Sim [5] in the linear case, with weaker assumptions on the underlying distribution of the uncertain parameter.

- For a variety of uncertainty sets of practical interest, we are able to compute in closed form or derive valid approximations of the robust complexity $\mathcal{Z}$, as summarized in Table 2. In particular, we provide the first results regarding polyhedra defined with a finite number of linear inequalities and the Minkowski sum of norm balls. From a geometric perspective, sum sets are very similar to intersection sets, while being computationally more tractable and leading to competitive probabilistic guarantees when properly scaled.
- We further improve those bounds a posteriori, i.e., given the resulting robust solution $\boldsymbol{x}$. In particular, in the linear case, we prove an a posteriori bound of the form

$$\mathbb{P}\left(\tilde{\boldsymbol{a}}^\top \boldsymbol{x} > b\right) \le \exp\left(-\frac{1}{2}\left[\frac{b-\bar{\boldsymbol{a}}^\top \boldsymbol{x}}{\|\boldsymbol{P}^\top \boldsymbol{x}\|_2}\right]^2\right),$$

and an analogous result in the case where the constraint is concave in the uncertainty, as summarized in Table 1. Besides being tighter, a posteriori bounds can notably be computed for any general convex uncertainty set.

The rest of the paper is structured as follows: Section 2 recalls equivalent reformulations for general uncertain constraints [4]. We formally state and prove the a posteriori and a priori bounds for general constraints and uncertainty sets in Section 3. Section 4 provides some closed form expressions or valid approximations for the robust complexity of a set, which drives a priori guarantees. Finally, we illustrate how our results can be applied in practice on a facility location and a portfolio optimization problem in Section 5.

Notations In the remainder of the paper, we use nonbold (x), lowercase bold ($\boldsymbol{x}$) and uppercase bold ($\boldsymbol{X}$) characters to denote scalars, vectors and matrices, respectively. Calligraphic characters such as $\mathcal{X}$ denote sets. We use a tilde symbol (e.g., $\tilde{x}$) to indicate a random variable. We let $\mathbf{e}$ denote the vector of all 1's, $\mathbf{0}$ denote the vector of all 0's and $\boldsymbol{e}_i$ the ith vector of the canonical basis, with dimension implied by the context. For any $p \in \mathbb{N}$, we define the ℓ_p-norm of $\boldsymbol{x} \in \mathbb{R}^n$ as $\|\boldsymbol{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$. For $p = \infty$, $\|\boldsymbol{x}\|_\infty := \max_i |x_i|$. Any function $f : \mathbb{R}^n \to \mathbb{R}$ can be defined on a subset of $\mathbb{R}^n$ only, called its domain and denoted $\mathrm{dom}(f)$. Similarly, f can be extended to the whole space $\mathbb{R}^n$ by setting $f(\boldsymbol{x}) = \infty$ if $\boldsymbol{x} \notin \mathrm{dom}(f)$. For any convex nonempty set $\mathcal{Z}$, its relative interior is defined and denoted $\mathrm{ri}(\mathcal{Z}) := \{\boldsymbol{x} \in \mathcal{Z} : \forall \boldsymbol{y} \in \mathcal{Z}\ \exists \lambda > 1, \boldsymbol{y} + \lambda(\boldsymbol{x} - \boldsymbol{y}) \in \mathcal{Z}\}$. We denote $\mathcal{S}_+^n$ the cone of $n \times n$ positive semi-definite matrices, $\mathcal{S}_+^n := \{\boldsymbol{A} \in \mathbb{R}^{n\times n} \ : \ \boldsymbol{A} \succeq 0\}$.

2 Robust counterparts of general uncertain constraint

In this section, we recall useful results from Ben-Tal et al. [4] about tractable reformulations of the robust constraint

$$f(\bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z}, \boldsymbol{x}) \leq 0,\ \forall \boldsymbol{z} \in \mathcal{Z}, \tag{5}$$

when the function $f(\boldsymbol{a}, \boldsymbol{x})$ is concave in the uncertain parameter $\boldsymbol{a}$.

2.1 Case when the constraint is linear

We first consider the case where f is linear, $f(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} - b$, and consider the robust constraint

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \boldsymbol{z}^\top \boldsymbol{P}^\top \boldsymbol{x} \leq b,\ \forall \boldsymbol{z} \in \mathcal{Z}.$$

Proposition 1 *The decision variable $\boldsymbol{x}$ satisfies*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \boldsymbol{z}^\top \boldsymbol{P}^\top \boldsymbol{x} \leq b,\ \forall \boldsymbol{z} \in \mathcal{Z},$$

if and only if it satisfies

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \delta^\star\left(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z}\right) \leq b,$$

where $\delta^\star(\boldsymbol{y}|\mathcal{Z}) := \sup_{\boldsymbol{z} \in \mathcal{Z}} \boldsymbol{y}^\top \boldsymbol{z}$ is the so-called support function *of $\mathcal{Z}$ [35, chapter 13].*

Proposition 1 uncovers the role played by the uncertainty set $\mathcal{Z}$ through its support function $\delta^\star(\cdot|\mathcal{Z})$. Though general, Proposition 1 is also a practical statement, for the support function can be computed for a wide range of uncertainty sets [see 4, Section 3]. We report some of these results in Table 3. Observe that whenever the support function is expressed as a minimization problem, the "min" operator can be omitted in the robust counterpart of Proposition 1, given the sense of the inequality.

Table 3: Examples of uncertainty sets and their associated support function.

Uncertainty region	Definition	Support function $\delta^\star(\boldsymbol{y}\|\mathcal{Z})$
Box	$\\|\boldsymbol{z}\\|_\infty \leq \rho$	$\rho\\|\boldsymbol{y}\\|_1$
Ball	$\\|\boldsymbol{z}\\|_2 \leq \rho$	$\rho\\|\boldsymbol{y}\\|_2$
Norm	$\\|\boldsymbol{z}\\| \leq \rho$	$\rho\\|\boldsymbol{y}\\|_\star$
Budget	$\\|\boldsymbol{z}\\|_\infty \leq 1$ $\\|\boldsymbol{z}\\|_1 \leq \rho$	$\min_{\boldsymbol{v}} \\|\boldsymbol{v}\\|_1 + \rho\\|\boldsymbol{y} - \boldsymbol{v}\\|_\infty$
Polyhedral	$\boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}$	$\min_{\boldsymbol{v} \geq \boldsymbol{0} : \boldsymbol{D}^\top \boldsymbol{v} = \boldsymbol{y}} \boldsymbol{d}^\top \boldsymbol{v}$
Intersection	$\mathcal{Z}_1 \cap \mathcal{Z}_2$	$\min_{\boldsymbol{v}} \delta^\star(\boldsymbol{v}\|\mathcal{Z}_1) + \delta^\star(\boldsymbol{y} - \boldsymbol{v}\|\mathcal{Z}_2)$
Minkowski sum	$\mathcal{Z}_1 + \mathcal{Z}_2$	$\delta^\star(\boldsymbol{y}\|\mathcal{Z}_1) + \delta^\star(\boldsymbol{y}\|\mathcal{Z}_2)$

Example 1 If $\mathcal{Z} = \{\boldsymbol{z} \ : \ \boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}\}$ is a polyhedron, then the support function of $\mathcal{Z}$ is given by

$$\delta^\star(\boldsymbol{y}|\mathcal{Z}) = \min_{\boldsymbol{v} \geq \boldsymbol{0} : \boldsymbol{D}^\top \boldsymbol{v} = \boldsymbol{y}} \boldsymbol{d}^\top \boldsymbol{v}.$$

According to Proposition 1, the robust linear constraint is hence equivalent to

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \min_{\boldsymbol{v} \geq \boldsymbol{0} : \boldsymbol{D}^\top \boldsymbol{v} = \boldsymbol{y}} \boldsymbol{d}^\top \boldsymbol{v} \leq b,$$

which in turn is equivalent to the existence of a feasible vector $\boldsymbol{v}$, $\boldsymbol{v} \geq \boldsymbol{0}$ and $\boldsymbol{D}^\top \boldsymbol{v} = \boldsymbol{y}$, satisfying $\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \boldsymbol{d}^\top \boldsymbol{v} \leq b$. Indeed, the inequality holds for some $\boldsymbol{v}$ if and only if it holds for the minimum.

2.2 Case when the constraint is concave in the uncertainty

We now consider the case where $f(\boldsymbol{a}, \boldsymbol{x})$ is a concave function in $\boldsymbol{a}$. Indeed, constraint (5) is equivalent to

$$\max_{\boldsymbol{z} \in \mathcal{Z}} f(\bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z}, \boldsymbol{x}) \leq 0,$$

and the inner maximization problem is well posed. In this case, Ben-Tal et al. [4] provide an equivalent reformulation under the technical assumption that $\bar{\boldsymbol{a}}$ is *regular*, that is when $\bar{\boldsymbol{a}}$ is within the relative interior of the domain of f, $\bar{\boldsymbol{a}} \in \mathrm{ri}(\mathrm{dom}(f(\cdot, \boldsymbol{x})))$ in short. From a high-level perspective, they require f to be properly defined in the vicinity of the nominal value $\bar{\boldsymbol{a}}$. So $\bar{\boldsymbol{a}}$ should not lie at the boundary of the domain of f. For the rest of the paper, we will make this assumption, when needed.

Proposition 2 *If $\bar{\boldsymbol{a}} \in ri(dom(f(\cdot, \boldsymbol{x})))$ and $f(\boldsymbol{a}, \boldsymbol{x})$ is concave in $\boldsymbol{a}$, then $\boldsymbol{x}$ satisfies*

$$f(\bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z}, \boldsymbol{x}) \leq 0, \ \forall \boldsymbol{z} \in \mathcal{Z},$$

if and only if there exists some vector $\boldsymbol{v}$ such that $(\boldsymbol{x}, \boldsymbol{v})$ satisfies

$$\bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star \left(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z}\right) - f_\star(\boldsymbol{v}, \boldsymbol{x}) \leq 0,$$

where $\delta^\star(\cdot|\mathcal{Z})$ is the support function *of $\mathcal{Z}$ and $f_\star(\boldsymbol{v}, \boldsymbol{x}) := \inf_{\boldsymbol{a}} \ \boldsymbol{a}^\top \boldsymbol{v} - f(\boldsymbol{a}, \boldsymbol{x})$.*

Remarkably, Proposition 2 provides a robust reformulation of the nonlinear constraints (5) where the terms involving $\mathcal{Z}$ are independent from those involving f. In this regard, the auxiliary variable $\boldsymbol{v}$ plays a critical role: $\boldsymbol{v}$ is linearly impacted by the uncertainty on $\boldsymbol{z}$ while being coupled with $\boldsymbol{x}$ through the conjugate function $f_\star$. In the special case where $f(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} - b$ is linear in $\boldsymbol{a}$, then

$$f_\star(\boldsymbol{v}, \boldsymbol{x}) = \begin{cases} b, & \text{if } \boldsymbol{v} = \boldsymbol{x}, \\ \infty, & \text{otherwise,} \end{cases}$$

and we recover Proposition 1.

Example 2 We consider a case where f is quadratic in $\boldsymbol{a}$ and linear in $\boldsymbol{x}$, that is $f(\boldsymbol{a}, \boldsymbol{x}) = -\frac{1}{2} \sum_i (\boldsymbol{a}^\top \boldsymbol{Q}_i \boldsymbol{a}) x_i$ for some positive semi-definite matrices $\boldsymbol{Q}_i$'s. Denoting $\boldsymbol{Q}(\boldsymbol{x}) := \sum_i x_i \boldsymbol{Q}_i$,

$$f_\star(\boldsymbol{v}, \boldsymbol{x}) = \inf_{\boldsymbol{a}} \ \boldsymbol{a}^\top \boldsymbol{v} + \frac{1}{2} \boldsymbol{a}^\top \boldsymbol{Q}(\boldsymbol{x}) \boldsymbol{a} = -\frac{1}{2} \boldsymbol{v} \boldsymbol{Q}(\boldsymbol{x})^{-1} \boldsymbol{v}.$$

Hence, the robust constraint becomes

$$\exists \boldsymbol{v}, \ \bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star \left(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z} \right) + \frac{1}{2} \boldsymbol{v} \boldsymbol{Q}(\boldsymbol{x})^{-1} \boldsymbol{v} \leq 0.$$

For tractability, it might be useful to observe that $f_\star(\boldsymbol{v}, \boldsymbol{x})$ can also be written [4, Lemma 6.2]

$$f_\star(\boldsymbol{v}, \boldsymbol{x}) = \sup_{\boldsymbol{s}^i : \sum_i \boldsymbol{s}^i = \boldsymbol{v}} \ -\sum_i \frac{1}{2x_i} \boldsymbol{s}^i \boldsymbol{Q}_i^{-1} \boldsymbol{s}^i,$$

and consider the equivalent reformulation

$$\exists \boldsymbol{v}, \boldsymbol{s}^i, \ \sum_i \boldsymbol{s}^i = \boldsymbol{v}, \ \text{and} \ \ \bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star \left(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z} \right) + \sum_i \frac{1}{2x_i} \boldsymbol{s}^i \boldsymbol{Q}_i^{-1} \boldsymbol{s}^i \leq 0,$$

which is conic quadratic representable.

Example 3 We consider the mean-variance portfolio problem with n assets

$$\min_{\boldsymbol{x} \geq 0 : \boldsymbol{e}^\top \boldsymbol{x} = 1} \ -\boldsymbol{\mu}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x},$$

where $\lambda > 0$ is a risk-aversion coefficient. In practice, the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the returns are estimated empirically, hence inherently noisy. To account for uncertainty, we write the nominal problem in epigraph formulation,

$$\min_{t, \boldsymbol{x} \geq 0 : \boldsymbol{e}^\top \boldsymbol{x} = 1} \ t \ \text{ s.t. } t \geq -\boldsymbol{\mu}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x},$$

and consider the robust counterpart of the constraint $f(\boldsymbol{a}, \boldsymbol{x}) \leq 0$ with $\boldsymbol{a} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^n \times \mathcal{S}_+^n$ and

$$f(\boldsymbol{a}, \boldsymbol{x}) = -t - \boldsymbol{\mu}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}.$$

Observe that f is quadratic in $\boldsymbol{x}$ and linear in $\boldsymbol{a}$.

Any element $\boldsymbol{a} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^n \times \mathcal{S}_+^n$ can be associated with a $(n + n^2)$-dimensional vector obtained by concatenating $\boldsymbol{\mu}$ and the columns of $\boldsymbol{\Sigma}$, denoted $\mathrm{vec}(\boldsymbol{a})$. As in Marandi et al. [32], we consider the uncertainty set

$$(\mathbb{R}^n \times \mathcal{S}_+^n) \cap \{ \boldsymbol{a} : \mathrm{vec}(\boldsymbol{a}) \in \mathcal{U} \},$$

where $\mathcal{U} \subset \mathbb{R}^{n+n^2}$ is of the desired form. For concision, we omit the vec$(\cdot)$ operator in the rest of the paper whenever the context is clear. In a similar vein, we will concisely refer to the uncertainty set above as $(\mathbb{R}^n \times \mathcal{S}_+^n) \cap \mathcal{U}$.

Proposition 2 applies to functions defined over $\boldsymbol{a} \in \mathbb{R}^m$. The dual variable $\boldsymbol{v}$ belongs to the same space. However, here, $\boldsymbol{a} \in \mathbb{R}^n \times \mathcal{S}_+^n$. Yet, Proposition 2 is still valid, provided that the additional variables $\boldsymbol{v}$ are taken in $\mathbb{R}^n \times \mathcal{S}_+^n$ as well. This generalization from $\mathbb{R}^m$ to $\mathbb{R}^n \times \mathcal{S}_+^n$ is justified by duality results for general inequalities [11, Chapter 5.9].

Consequently, we decompose the auxiliary variable $\boldsymbol{v}$ into $\boldsymbol{v} = (\boldsymbol{w}, \boldsymbol{W}) \in \mathbb{R}^n \times \mathcal{S}_+^n$. Denoting $\langle \cdot, \cdot \rangle$ the Euclidean inner-product of matrices, we have:

$$\begin{aligned}
\bar{\boldsymbol{a}}^\top \boldsymbol{v} &= \bar{\boldsymbol{\mu}}^\top \boldsymbol{w} + \langle \bar{\boldsymbol{\Sigma}}, \boldsymbol{W} \rangle, \\
f_\star(\boldsymbol{v}, \boldsymbol{x}) &= \inf_{\boldsymbol{\mu}, \boldsymbol{\Sigma} \succeq 0} \; \boldsymbol{\mu}^\top \boldsymbol{w} + \langle \boldsymbol{\Sigma}, \boldsymbol{W} \rangle + t + \boldsymbol{\mu}^\top \boldsymbol{x} - \lambda \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}, \\
&= t + \inf_{\boldsymbol{\mu}} \; \boldsymbol{\mu}^\top (\boldsymbol{w} + \boldsymbol{x}) \; + \inf_{\boldsymbol{\Sigma} \succeq 0} \langle \boldsymbol{\Sigma}, \boldsymbol{W} - \lambda \boldsymbol{x} \boldsymbol{x}^\top \rangle, \\
&= \begin{cases} t, & \text{if } \boldsymbol{w} = -\boldsymbol{x} \text{ and } \boldsymbol{W} - \lambda \boldsymbol{x}\boldsymbol{x}^\top \succeq 0, \\ \infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

So, according to Proposition 2, the robust constraint is equivalent to

$$\exists \boldsymbol{W} \succeq 0 \text{ s.t. } \boldsymbol{W} \succeq \lambda \boldsymbol{x}\boldsymbol{x}^\top \text{ and } t \geq -\bar{\boldsymbol{\mu}}^\top \boldsymbol{x} + \langle \bar{\boldsymbol{\Sigma}}, \boldsymbol{W} \rangle + \delta^\star(\boldsymbol{P}^\top \text{vec}(-\boldsymbol{x}, \boldsymbol{W}) | \mathcal{Z}).$$

All in all, we can solve the robust mean-variance portfolio problem as the following semi-definite optimization problem

$$\begin{aligned}
\min_{t, \boldsymbol{x} \geq 0 : \boldsymbol{e}^\top \boldsymbol{x} = 1, \boldsymbol{W} \succeq 0} \; t \;\; \text{s.t.} \; & \begin{pmatrix} \boldsymbol{W} & \boldsymbol{x} \\ \boldsymbol{x}^\top & 1/\lambda \end{pmatrix} \succeq 0, \\
& t \geq -\bar{\boldsymbol{\mu}}^\top \boldsymbol{x} + \langle \bar{\boldsymbol{\Sigma}}, \boldsymbol{W} \rangle + \delta^\star(\boldsymbol{P}^\top \text{vec}(-\boldsymbol{x}, \boldsymbol{W}) | \mathcal{Z}).
\end{aligned}$$

We refer the reader to Ben-Tal et al. [4] Section 4, Table 2 for more examples.

3 A priori and a posteriori probabilistic guarantees

Most of the comparable results in the literature are valid under some light-tail assumption on the random vector $\tilde{\boldsymbol{z}}$, in particular that $\tilde{\boldsymbol{z}}$ is bounded almost surely or follows a Gaussian distribution. All our results in this section hold under the weaker assumption that the random vector $\tilde{\boldsymbol{z}}$ is sub-Gaussian.

Definition 1 [Definition 1.2 in 34] A random variable $\tilde{z} \in \mathbb{R}$ is said to be *sub-Gaussian* with parameter σ^2, denoted $\tilde{z} \sim \text{subG}\left(\sigma^2\right)$, if $\tilde{z}$ is centered, i.e., $\mathbb{E}[\tilde{z}] = 0$, and for all $s \in \mathbb{R}$,

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

Naturally, centered Gaussian random variables are also sub-Gaussian. Bounded random variables are also a special case of sub-Gaussian random variables, as a consequence of Hoeffding's inequality (see Appendix A, Lemma 1). The parameter σ^2 is usually referred to as a *variance proxy*, for in the case where $\tilde{z}$ is normally distributed, the inequality above holds with $\sigma^2 = \text{Var}(\tilde{z})$.

For the rest of the paper, we will make the following assumption on the randomness $\tilde{\boldsymbol{z}}$:

Assumption 1 *We assume that the coordinates of the uncertain parameter $\tilde{\boldsymbol{z}} \in \mathbb{R}^L$ are L independent sub-Gaussian random variables with variance proxy* 1.

Assumption 1 holds in particular if the coordinates of $\tilde{\boldsymbol{z}}$ are (a) L independent Gaussian random variables with mean 0 and variance 1, or (b) L independent centered random variables in $[-1,1]$, which is a commonly made assumption in the RO literature [2, 5]. We discuss alternative assumptions in Section 3.3 and report some useful properties of sub-Gaussian random variables in Appendix A.

3.1 Case when the constraint is linear

In this section, we consider the linear case where $f(\boldsymbol{a},\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} - b$. We first prove a solution-specific probabilistic guarantee:

Theorem 1 *Under Assumption 1, for any $\boldsymbol{x} \in \mathcal{X}$ satisfying*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z}) \leq b,$$

we have

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \exp\left(-\frac{(b-\bar{\boldsymbol{a}}^\top \boldsymbol{x})^2}{2\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2}\right) \leq \exp\left(-\frac{\delta^\star(\boldsymbol{P}^\top \boldsymbol{x}|\mathcal{Z})^2}{2\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2}\right).$$

Proof Consider a robust solution $\boldsymbol{x} \in \mathcal{X}$. We have

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) = \mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b - \bar{\boldsymbol{a}}^\top \boldsymbol{x}\right).$$

Let us denote $t := b - \bar{\boldsymbol{a}}^\top \boldsymbol{x} \geq \delta^\star(\boldsymbol{P}^\top \boldsymbol{x}|\mathcal{Z})$. If $\boldsymbol{P}^\top \boldsymbol{x} \neq 0$, $\delta^\star(\boldsymbol{P}^\top \boldsymbol{x}|\mathcal{Z}) > 0$ since $\mathcal{Z}$ is full dimensional. Under Assumption 1, $\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x}$ is sub-Gaussian with variance proxy $\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2$ (see Appendix A, Proposition 11). Hence, we have the following tail bound (Appendix A, Proposition 10)

$$\mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > t\right) \leq \exp\left(-\frac{t^2}{2\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2}\right).$$

In the case where $\boldsymbol{P}^\top \boldsymbol{x} = 0$, we have $\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} = \bar{\boldsymbol{a}}^\top \boldsymbol{x} \leq b$ for all $\tilde{\boldsymbol{z}}$, so that the probability of constraint violation is 0 and our bound holds with the convention $\frac{0}{0} = \infty$. □

The probabilistic guarantee of constraint violation in Theorem 1 depends on the specific solution $\boldsymbol{x}$. As a result, this bound cannot be used a priori to scale the size of the uncertainty set $\mathcal{Z}$ according to some target probabilistic guarantee. Yet, given a robust solution $\boldsymbol{x}$, the bound only involves quantities which can be easily computed and can provide a useful a posteriori guarantee, as we discuss in Section 5. From a practical standpoint, the bound involving the slack term $b - \bar{\boldsymbol{a}}^\top \boldsymbol{x}$ is systematically tighter, hence preferable. Yet, from a theoretical perspective, it is very satisfying to elicit how the uncertainty set impacts the probabilistic guarantee, through its support function at $\boldsymbol{P}^\top \boldsymbol{x}$, which also appears in the reformulation of the robust constraint. The latter bound will be instrumental in deriving set-specific but solution-independent a posteriori bound.

Remark 1 Theorem 1 can be used to obtain a robust solution $\boldsymbol{x}$ which would violate the uncertain constraint with probability at most ε by adding the second-order cone constraint

$$b - \bar{\boldsymbol{a}}^\top \boldsymbol{x} \geq \sqrt{2 \ln\left(\frac{1}{\varepsilon}\right)} \|\boldsymbol{P}^\top \boldsymbol{x}\|_2,$$

to the original optimization problem. This corresponds to the robust counterpart of the robust constraint with the ball uncertainty set, $\mathcal{Z} = \left\{ \boldsymbol{z} \ : \ \|\boldsymbol{z}\|_2 \leq \sqrt{2\ln\left(\frac{1}{\varepsilon}\right)} \right\}$.

Remark 2 The first inequality

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \exp\left(-\frac{(b - \bar{\boldsymbol{a}}^\top \boldsymbol{x})^2}{2\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2}\right)$$

is valid as long as $\boldsymbol{x}$ satisfies $\bar{\boldsymbol{a}}^\top \boldsymbol{x} < b$.

Theorem 1 can be extended to give a priori probabilistic guarantee which does not depend on a specific solution $\boldsymbol{x}$. To do so, we first define the *robust complexity* of a set $\mathcal{Z}$.

Definition 2 For any set $\mathcal{Z} \subseteq \mathbb{R}^L$, we define the *robust complexity* of $\mathcal{Z}$, and denote $\rho(\mathcal{Z})$, the quantity

$$\rho(\mathcal{Z}) := \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \delta^\star(\boldsymbol{y}|\mathcal{Z}) = \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \max_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{z}^\top \boldsymbol{y}.$$

In statistics, the Rademacher complexity of a set $\mathcal{Z}$ is defined as $\mathcal{R}(\mathcal{Z}) := \mathbb{E}_{\tilde{\boldsymbol{y}}}\left[\sup_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{z}^\top \tilde{\boldsymbol{y}}\right]$, where the coordinates of $\tilde{\boldsymbol{y}}$ are independently drawn from a Rademacher distribution, i.e., $\mathbb{P}(\tilde{y}_j = \pm 1) = 1/2$ [see 38, Example 2.25]. This quantity describes the size of the set $\mathcal{Z}$ and drives the so-called uniform law of large numbers [see 38, Chapters 4 and 5]. In this regard, our proposed complexity metric $\rho(\mathcal{Z})$ is a robust analog to the Rademacher complexity, replacing expectation by worst-case value. Indeed, for any Rademacher vector $\tilde{\boldsymbol{y}}$, we have $\|\tilde{\boldsymbol{y}}\|_2 = \sqrt{L}$ so that $\sqrt{L}\rho(\mathcal{Z}) \leq \mathcal{R}(\mathcal{Z})$.

Another measure of complexity studied in high-dimensional statistics is the Gaussian complexity, defined as $\mathcal{G}(\mathcal{Z}) := \mathbb{E}_{\tilde{\boldsymbol{y}}}\left[\sup_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{z}^\top \tilde{\boldsymbol{y}}\right]$, where the coordinates of $\tilde{\boldsymbol{y}}$ are independent Gaussian random variables with mean 0 and variance 1 [see 38, Example 2.30 and Chapter 5]. Since $\mathcal{R}(\mathcal{Z}) \leq \sqrt{\frac{\pi}{2}}\mathcal{G}(\mathcal{Z})$ [see 38, Exercise 5.5], we have $\rho(\mathcal{Z}) \leq \sqrt{\frac{\pi}{2L}}\mathcal{G}(\mathcal{Z})$. Alternatively, by decomposing $\tilde{\boldsymbol{y}}$ into $\tilde{r}\tilde{\boldsymbol{u}}$ where $\tilde{\boldsymbol{u}}$ is a unit vector uniformely distributed over the unit sphere and $\tilde{r}$ an independent scaling factor, we obtain $\mathcal{G}(\mathcal{Z}) \geq \rho(\mathcal{Z})\mathbb{E}[\tilde{r}]$, where $\mathbb{E}[\tilde{r}]$ is the expectation of a chi distribution with parameter L, i.e., $\mathbb{E}[\tilde{r}] = \sqrt{2}\Gamma((L+1)/2)/\Gamma(L/2)$ with $\Gamma(\cdot)$ denoting Euler's Gamma function. For large L, Stirling's approximation yields $\mathbb{E}[\tilde{r}] \sim \sqrt{L}$, which is tighter than the previous bound.

As for the Rademacher or Gaussian complexity, the robust complexity is homogeneous with respect to set inflation.

Proposition 3 *For any set $\mathcal{Z}$ and $\alpha > 0$, $\rho(\alpha\mathcal{Z}) = \alpha\rho(\mathcal{Z})$.*

Proof For any $\alpha > 0$, $\rho(\alpha\mathcal{Z}) = \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \max_{\boldsymbol{z}\in\alpha\mathcal{Z}} \boldsymbol{y}^\top \boldsymbol{z} = \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \max_{\boldsymbol{z}'\in\mathcal{Z}} \alpha\, \boldsymbol{y}^\top \boldsymbol{z}' = \alpha\rho(\boldsymbol{z})$. □

In addition, the robust complexity is invariant by orthogonal scaling, namely the following equality holds:

Proposition 4 *For any orthogonal matrix $\boldsymbol{U}$ and any set $\mathcal{Z}$, $\rho(\boldsymbol{U}\mathcal{Z}) = \rho(\mathcal{Z})$, with $\boldsymbol{U}\mathcal{Z} := \{\boldsymbol{U}\boldsymbol{z} : \boldsymbol{z} \in \mathcal{Z}\}$.*

Proof The matrix $\boldsymbol{U}$ being orthogonal, the change of variable $\boldsymbol{y} = \boldsymbol{U}\boldsymbol{y}'$ satisfies $\|\boldsymbol{y}\|_2^2 = \boldsymbol{y}'^\top \boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{y}' = \|\boldsymbol{y}'\|_2^2$, and $\rho(\mathcal{Z}) = \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \max_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{z}^\top \boldsymbol{y} = \min_{\boldsymbol{y}':\|\boldsymbol{y}'\|_2=1} \max_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{z}^\top \boldsymbol{U}^\top \boldsymbol{y}' = \min_{\boldsymbol{y}':\|\boldsymbol{y}'\|_2=1} \max_{\boldsymbol{z}'\in\boldsymbol{U}\mathcal{Z}} \boldsymbol{z}'^\top \boldsymbol{y}' = \rho(\boldsymbol{U}\mathcal{Z})$. □

Note that this property is not satisfied by the Rademacher complexity since the transformed random variable $\tilde{\boldsymbol{y}}' = \boldsymbol{U}^{-1}\tilde{\boldsymbol{y}}$ is not a Rademacher random variable. The Gaussian complexity, for which $\tilde{\boldsymbol{y}}$ is a normal random vector with independent standard coordinates, on the other hand, is also invariant by orthogonal scaling.

Finally, the robust complexity is *the* key quantity controlling a priori probabilistic guarantees.

Corollary 1 *Under Assumption 1, for any* $\boldsymbol{x} \in \mathcal{X}$ *satisfying*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z}) \leq b,$$

we have

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \exp\left(-\tfrac{1}{2}\rho\left(\mathcal{Z}\right)^2\right).$$

In Section 4, we provide explicit analytic expressions and valid lower bounds on $\rho(\mathcal{Z})$ for uncertainty sets found in the literature such as the budget [5] and the box-ellipsoidal [2] uncertainty sets.

Proof Taking the worst over all feasible $\boldsymbol{x}$ in the right hand-side of Theorem 1 yields

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \exp\left(-\frac{1}{2} \min_{x \in \mathcal{X}} \frac{\delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z})^2}{\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2}\right).$$

Then,

$$\min_{x \in \mathcal{X}} \frac{\delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z})^2}{\|\boldsymbol{P}^\top \boldsymbol{x}\|_2^2} = \left[\min_{x \in \mathcal{X}} \delta^\star\left(\left.\frac{\boldsymbol{P}^\top \boldsymbol{x}}{\|\boldsymbol{P}^\top \boldsymbol{x}\|_2}\right| \mathcal{Z}\right)\right]^2 \geq \left[\min_{\boldsymbol{y}: \|y\|_2 = 1} \delta^\star\left(\boldsymbol{y} | \mathcal{Z}\right)\right]^2. \qquad \square$$

3.2 Case when the constraint is concave in the uncertainty

We now state the analogue of Theorem 1 for the case where the constraint depends on the uncertainty in a concave manner as in Eq. (5).

Theorem 2 *Under Assumption 1, for any* $\boldsymbol{x} \in \mathcal{X}$ *and* $\boldsymbol{v}$ *satisfying*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z}) - f_\star(\boldsymbol{v}, \boldsymbol{x}) \leq 0,$$

we have

$$\mathbb{P}\left(f(\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}}, \boldsymbol{x}) > 0\right) \leq \exp\left(-\frac{(f_\star(\boldsymbol{v}, \boldsymbol{x}) - \bar{\boldsymbol{a}}^\top \boldsymbol{v})^2}{2\|\boldsymbol{P}^\top \boldsymbol{v}\|_2^2}\right) \leq \exp\left(-\frac{\delta^\star(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z})^2}{2\|\boldsymbol{P}^\top \boldsymbol{v}\|_2^2}\right).$$

Before deriving a formal proof of Theorem 2, we graphically explain the intuition behind the result in dimension $L = 2$, in Figure 1. Assume $(\tilde{z}_1, \tilde{z}_2) \in [-1, 1]^2$ (black squared box), which is a special case of Assumption 1. The shaded blue region corresponds to the uncertainty set $\mathcal{Z}$. Let $\boldsymbol{x}$ be a robust solution. It induces a constraint on $\boldsymbol{z}$, $f(\bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z}, \boldsymbol{x}) \leq 0$ which is satisfied by all $\boldsymbol{z} \in \mathcal{Z}$. By concavity, the region $\{\boldsymbol{z} : f(\bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z}, \boldsymbol{x}) > 0\}$ admits a supporting hyperplane (in red) and is contained within a half-space. Theorem 2 provides an explicit description of this halfspace, $\{\boldsymbol{z} : \bar{\boldsymbol{a}}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{P}\boldsymbol{z} - f_\star(\boldsymbol{v}, \boldsymbol{x}) > 0\}$.

Proof Since

$$\bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z}) - f_\star(\boldsymbol{v}, \boldsymbol{x}) \leq 0,$$

we can apply Theorem 1 (with $b = f_\star(\boldsymbol{v}, \boldsymbol{x})$) and get

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{v} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{v} - f_\star(\boldsymbol{v}, \boldsymbol{x}) > 0\right) \leq \exp\left(-\frac{(f_\star(\boldsymbol{v}, \boldsymbol{x}) - \bar{\boldsymbol{a}}^\top \boldsymbol{v})^2}{2\|\boldsymbol{P}^\top \boldsymbol{v}\|_2^2}\right).$$

By definition of the conjugate,

$$\forall \boldsymbol{a}, \; f_\star(\boldsymbol{v}, \boldsymbol{x}) + f(\boldsymbol{a}, \boldsymbol{x}) \leq \boldsymbol{a}^\top \boldsymbol{v},$$

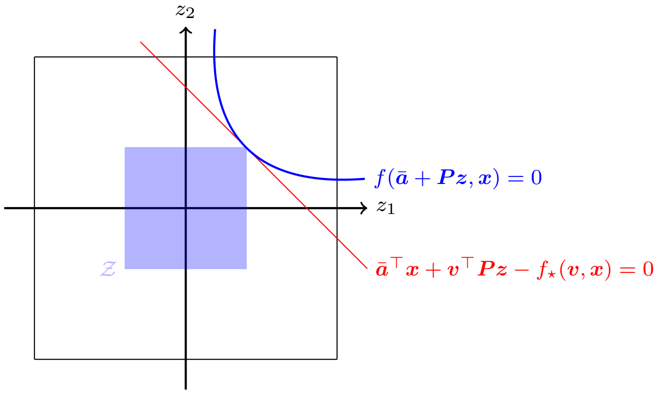


Fig. 1: Geometrical proof of Theorem 2 in dimension $L = 2$.

which in turns leads to

$$f(\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}}, \boldsymbol{x}) > 0 \implies (\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}})^\top \boldsymbol{v} - f_\star(\boldsymbol{v}, \boldsymbol{x}) \geq f(\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}}, \boldsymbol{x}) > 0.$$

All in all, we have

$$\begin{aligned} \mathbb{P}\left(f(\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}}, \boldsymbol{x}) > 0\right) &\leq \mathbb{P}\left((\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}})^\top \boldsymbol{v} - f_\star(\boldsymbol{v}, \boldsymbol{x}) > 0\right) \\ &\leq \exp\left(-\frac{(f_\star(\boldsymbol{v}, \boldsymbol{x}) - \bar{\boldsymbol{a}}^\top \boldsymbol{v})^2}{2\|\boldsymbol{P}^\top \boldsymbol{v}\|_2^2}\right), \end{aligned}$$

which concludes the proof. □

Theorem 2 displays the exact same probabilistic guarantee as in the linear case except that the bound now involves the extra variable $\boldsymbol{v}$ instead of $\boldsymbol{x}$. Recall that in the special case where f is linear, the conjugate $f_\star$ enforces $\boldsymbol{v} = \boldsymbol{x}$ and we recover Theorem 1. In the general case, the variable $\boldsymbol{v}$ is introduced to express the robust constraint in a tractable manner, so solving the robust optimization problem provides a vector $\boldsymbol{v}$ alongside a vector $\boldsymbol{x}$, and our bound can be computed a posteriori. Taking the worst-case over all potential vectors $\boldsymbol{v}$, we get the exact same *a priori* bounds as in the linear case.

Corollary 2 *Under Assumption 1, for any $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{v}$ satisfying*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{v} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{v} | \mathcal{Z}) - f_\star(\boldsymbol{v}, \boldsymbol{x}) \leq 0,$$

we have

$$\mathbb{P}\left(f(\bar{\boldsymbol{a}} + \boldsymbol{P}\tilde{\boldsymbol{z}}, \boldsymbol{x}) > 0\right) \leq \exp\left(-\tfrac{1}{2}\rho(\mathcal{Z})^2\right).$$

3.3 Alternative assumptions on the uncertain parameter $\tilde{\boldsymbol{z}}$

So far, we assumed that the coordinates of $\tilde{\boldsymbol{z}}$ were independent sub-Gaussian random variables with variance proxy 1, leading to the tail bound

$$\mathbb{P}\left(\tilde{z}_j > t\right) \leq e^{-\frac{t^2}{2\sigma^2}},$$

Table 4: Summary of relevant assumptions on $\tilde{\boldsymbol{z}}$ found in the literature and the corresponding variance proxy. Proofs for the last three lines can be found in Ben-Tal et al. [3, Chapter 2.4.]. If $\tilde{\boldsymbol{z}}$ has variance proxy σ^2 then the rate $1/2$ in the probabilistic guarantee should be replaced by $1/2\sigma^2$.

Assumption on $\tilde{z}$	Variance proxy
$\mathbb{E}[\tilde{z}] = 0$ $\lvert\tilde{z}\rvert \leq 1,$ a.s.	1
$\tilde{z} \sim \mathcal{N}(0,1)$	1
$\mathbb{E}[\tilde{z}]$ symmetric w.r.t. 0 $\mathbb{E}[\tilde{z}]$ unimodal w.r.t. 0 $\lvert\tilde{z}\rvert \leq 1,$ a.s.	$1/3$
$\mathbb{E}[\tilde{z}]$ symmetric w.r.t. 0 $\lvert\tilde{z}\rvert \leq 1,$ a.s. $\mathrm{Var}(\tilde{z}) \leq \nu^2$	$\min_{c \geq 0} \left\{ c \, : \, \forall t, \; c^2 t^2 \geq 2 \ln(\nu^2 \cosh(t) + 1 - \nu^2) \right\} \leq 1$
$\mathbb{E}[\tilde{z}]$ symmetric w.r.t. 0 $\mathbb{E}[\tilde{z}]$ unimodal w.r.t. 0 $\lvert\tilde{z}\rvert \leq 1,$ a.s. $\mathrm{Var}(\tilde{z}) \leq \nu^2 \leq 1/3$	$\min_{c \geq 0} \left\{ c \, : \, \forall t, \; c^2 t^2 \geq 2 \ln(3\nu^2 \sinh(t)/t + 1 - 3\nu^2) \right\} \leq 1$

with $\sigma^2 = 1$. We recall main properties and useful concentration inequalities for sub-Gaussian random variables in Appendix A. More generally, if the coordinates of $\tilde{\boldsymbol{z}}$ are independent sub-Gaussian random variables with variance proxy σ^2, then our results hold after properly dividing the exponential rate by a factor σ^2. As summarized in Table 4, common assumptions on the uncertainty parameter $\tilde{\boldsymbol{z}}$ found in the literature can be all seen as assuming sub-Gaussian random variables for some well-chosen variance proxy. For instance,

- Normal random variables with mean 0 and variance 1 are sub-Gaussian random variables with variance proxy 1,
- Bounded random variables in $[-1, 1]$ with mean 0 are sub-Gaussian random variables with variance proxy 1,
- Bounded random variables in $[-1, 1]$ which are symmetric and unimodal with respect to 0 are sub-Gaussian random variables with variance proxy $1/3$.

We now discuss how additional assumptions on the true distribution of $\tilde{\boldsymbol{z}}$ can be incorporated into our proof techniques to derive valid bounds. From a high-level perspective, the proofs of Theorems 1 and 2 are essentially a tail-bound $\mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > t\right)$ for some $t \geq \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z}) > 0$.

Independence: A key assumption is the independence of the coordinates of $\tilde{\boldsymbol{z}}$. Under this assumption, a standard Chernoff's bound technique leads to

$$\mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > t\right) \leq \min_{\theta > 0} \exp\left(-\theta t + \sum_{k=1}^{L} \log \mathbb{E}[e^{\theta \boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{e}_j \tilde{z}_j}]\right).$$

Then, further assumptions on $\tilde{\boldsymbol{z}}$ enables to bound the logarithm of the moment generating function and derive computable bounds. This is the approach taken in Ben-Tal et al. [3, chapter 2] for safe approximation of general scalar chance constraint. Again, we emphasize that while we assume independence of the coordinates of $\tilde{\boldsymbol{z}}$, we do not assume the coordinates of the uncertain vector $\tilde{\boldsymbol{a}}$ to be independent.

Bounded symmetric distribution: Under the assumption that the coordinates of $\tilde{\boldsymbol{z}}$ are L independent, symmetrically distributed random variables in $[-1,1]$, Bertsimas and Sim [5] prove that

$$\mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > t\right) \leq \frac{1}{2^L}\left[(1-\mu)\sum_{\ell=\lfloor\nu\rfloor}^{L}\binom{L}{\ell} + \mu \sum_{\ell=\lfloor\nu\rfloor+1}^{L}\binom{L}{\ell}\right],$$

with $\nu = \left(\frac{t}{\|\boldsymbol{P}^\top\boldsymbol{x}\|_2} + L\right)/2$, $\mu = \nu - \lfloor\nu\rfloor$. In particular, they show that this bound is tight for some particular distribution on $\tilde{\boldsymbol{z}}$. However, this bound might be hard to compute numerically and that provide tractable safe approximations of the upper-bound that match our results.

Bound on the covariance matrix $\mathbb{E}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^\top]$*:* If we assume as in Bertsimas et al. [7] that the coordinates of $\tilde{\boldsymbol{z}}$ are L random variables with zero mean and covariance matrix bounded by some positive semi-definite matrix $\boldsymbol{\Sigma} \succeq 0$, i.e., $\mathbb{E}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^\top] \preceq \boldsymbol{\Sigma}$, then Chebyshev's inequality yields

$$\mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > t\right) \leq \frac{\mathrm{Var}(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x})}{t^2} \leq \frac{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{P}^\top\boldsymbol{x}\|^2}{t^2}.$$

This tail bound could be used to provide both a priori and a posteriori probabilistic guarantees under this more general set of assumptions. The dependency in the robust complexity of the uncertainty set $\mathcal{Z}$, however, would be polynomial instead of exponential. We provide such bounds in Appendix B.

4 The robust complexity of a set

In this section, we derive explicit lower bounds for the robust complexity $\rho(\mathcal{Z})$ for particular uncertainty sets $\mathcal{Z}$ and recover existing results in the literature. We also discuss how those bounds should guide modeling in practice.

4.1 Norm-ball uncertainty sets

In practice, $\mathcal{Z}$ is often chosen as an ℓ_p ball for which one can compute its robust complexity explicitly.

Proposition 5 $\mathcal{Z} = \left\{\boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{z}\|_p \leq \Gamma\right\}$ *has a robust complexity of*

$$\rho(\mathcal{Z}) = \Gamma\kappa(p), \text{ with } \kappa(p) := \begin{cases} 1, & \text{if } p \geq 2, \\ L^{1/2 - 1/p}, & \text{if } p \leq 2. \end{cases}$$

Proof Since the support function for an ℓ_p-norm ball is given by its dual norm, the ℓ_q-norm with $q \in [1,\infty]$ satisfying $1/p + 1/q = 1$, we derive

$$\min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \delta^\star(\boldsymbol{y}|\mathcal{Z}) = \Gamma \min_{\boldsymbol{y}:\|\boldsymbol{y}\|_2=1} \|\boldsymbol{y}\|_q.$$

If $q \leq 2$, $1 = \|\boldsymbol{y}\|_2 \leq \|\boldsymbol{y}\|_q$, which is tight for $\boldsymbol{y} = \boldsymbol{e}_i$ for some $i \in \{1,\ldots,L\}$. If $q \geq 2$, Hölder's inequality yields $1 = \|\boldsymbol{y}\|_2 \leq L^{1/2-1/q}\|\boldsymbol{y}\|_q$, which is tight for $\boldsymbol{y} = \sqrt{L}\boldsymbol{e}$. Hence, the result. □

This result generalizes results obtained in the literature. In particular, if the 2-norm is used, the bound is notably independent of the dimension L, and a priori and a posteriori bounds match. As intuition suggests, the bound monotonically decreases with Γ, that is, the bigger the uncertainty set, the smaller the probability of constraint violation. Equivalently, in order for the constraint to be violated with probability at most ε, it suffices to take $\Gamma \geq \frac{1}{\kappa(p)} \sqrt{2 \ln(1/\varepsilon)}$.

Although its generality can be appealing, Proposition 5 might be quite weak in presence of specific problem structure. For instance, for $p = \infty$ and $\Gamma = 1$, the upper-bound is $e^{-1/2} \approx 61\%$, while constraint violation happens with probability zero given Assumption 1.

Proposition 5 applies to norm balls which are isotropic, i.e., which are invariant by permutation of the coordinates. For uncertainty sets which weight each coordinate z_i by a specific weight $\lambda_i > 0$, we can prove the following extension.

Proposition 6 *Let $\boldsymbol{\Lambda} = Diag(\boldsymbol{\lambda})$ be a diagonal matrix with $\boldsymbol{\lambda} > \mathbf{0}$ and let $p \in [1, \infty]$. Then the set $\mathcal{Z} = \left\{ \boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{\Lambda z}\|_p = \left(\sum_i |\lambda_i z_i|^p\right)^{1/p} \leq 1 \right\}$ has a robust complexity of*

$$\rho(\mathcal{Z}) = 1/\|\boldsymbol{\lambda}\|_\infty, \qquad \text{if } p \geq 2,$$

$$\rho(\mathcal{Z}) \geq \left(\sum_{i=1}^L \lambda_i^{1/(1/p - 1/2)} \right)^{1/2 - 1/p}, \qquad \text{if } p \leq 2.$$

In particular, if $\lambda_i = 1/\Gamma$, we recover Proposition 5.

Proof By a change of variable, we have

$$\delta^\star(\boldsymbol{y}|\mathcal{Z}) = \max_{\boldsymbol{z}' : \|\boldsymbol{z}'\|_p \leq 1} \boldsymbol{y}^\top \Lambda^{-1} \boldsymbol{z}' = \|\Lambda^{-1} \boldsymbol{y}\|_q,$$

with $1/p + 1/q = 1$. If $q \leq 2$ (i.e., $p \geq 2$),

$$1 = \|\boldsymbol{y}\|_2 = \sqrt{\sum_{i=1}^L \lambda_i^2 \left(\frac{y_i}{\lambda_i}\right)^2} \leq \|\boldsymbol{\lambda}\|_\infty \|\boldsymbol{\Lambda}^{-1} \boldsymbol{y}\|_2 \leq \|\boldsymbol{\lambda}\|_\infty \|\boldsymbol{\Lambda}^{-1} \boldsymbol{y}\|_q,$$

which is tight for $\boldsymbol{y} = \boldsymbol{e}_i$ with $i \in \arg\max_i \ \lambda_i$. If $q > 2$ (i.e., $p < 2$), , Hölder's inequality yields

$$1 = \|\boldsymbol{y}\|_2 = \sqrt{\sum_{i=1}^L \lambda_i^2 \left(\frac{y_i}{\lambda_i}\right)^2} \leq \left(\sum_{i=1}^L \lambda_i^{1/\alpha} \right)^\alpha \|\boldsymbol{\Lambda}^{-1} \boldsymbol{y}\|_q,$$

with $\alpha = \frac{1}{2}\left(1 - \frac{2}{q}\right) = \frac{1}{p} - \frac{1}{2}$. □

4.2 Intersection of norm balls

Uncertainty sets defined as the intersection of norm-balls have attracted a lot of attention, because of their tractability and reduced conservatism. For such sets, one can compute an explicit a priori probabilistic guarantee.

Proposition 7 *Let $\mathcal{Z}_i$, $i = 1, 2$, be two uncertainty sets such that $ri(\mathcal{Z}_1) \cap ri(\mathcal{Z}_2) \neq \emptyset$.*

(a) The robust complexity of $\mathcal{Z}_1 \cap \mathcal{Z}_2$ satisfies $\rho(\mathcal{Z}_1 \cap \mathcal{Z}_2) \geq \min_{i=1,2} \rho(\mathcal{Z}_i)$.
(b) If, in addition to Assumption 1, $\tilde{\boldsymbol{z}} \in \mathcal{Z}_1$ almost surely, then the probabilistic guarantee is driven by $\rho(\mathcal{Z}_2)$.

Proposition 7(a) is non trivial. Since $\mathcal{Z}_1 \cap \mathcal{Z}_2 \subseteq \mathcal{Z}_i$, $i = 1, 2$, the constraint violation probability induced by $\mathcal{Z}_1 \cap \mathcal{Z}_2$ can only be worse than the ones induced by $\mathcal{Z}_1$ and $\mathcal{Z}_2$ alone. Yet, Proposition 7(a) states that, as far as upper-bounds are concerned, $\mathcal{Z}_1 \cap \mathcal{Z}_2$ is no worse than $\mathcal{Z}_1$ and $\mathcal{Z}_2$ separately. Proposition 7(b) improves upon Proposition 7(a) under the additional assumption that $\tilde{z} \in \mathcal{Z}_1$ almost surely, stating that $\mathcal{Z}_2$ alone controls the upper bound for constraint violation.

Proof (a) If $\mathcal{Z} = \mathcal{Z}_1 \cap \mathcal{Z}_2$ with $\mathrm{ri}(\mathcal{Z}_1) \cap \mathrm{ri}(\mathcal{Z}_2) \neq \emptyset$, we have [4, Lemma 6.4] $\delta^\star(\boldsymbol{y}|\mathcal{Z}) = \min_{\boldsymbol{v}} \delta^\star(\boldsymbol{v}|\mathcal{Z}_1) + \delta^\star(\boldsymbol{y} - \boldsymbol{v}|\mathcal{Z}_2)$. In addition, by definition of the robust complexity $\delta^\star(\cdot|\mathcal{Z}_i) \geq \rho(\mathcal{Z}_i)\|\cdot\|_2$ so that

$$
\begin{aligned}
\delta^\star(\boldsymbol{y}|\mathcal{Z}) &\geq \min_{\boldsymbol{v}} \; \rho(\mathcal{Z}_1)\|\boldsymbol{v}\|_2 + \rho(\mathcal{Z}_2)\|\boldsymbol{y} - \boldsymbol{v}\|_2 \\
&\geq \min_{i=1,2} \; \rho(\mathcal{Z}_i) \; \min_{\boldsymbol{v}} \; [\|\boldsymbol{v}\|_2 + \|\boldsymbol{y} - \boldsymbol{v}\|_2] \, , \\
&\geq \min_{i=1,2} \; \rho(\mathcal{Z}_i) \; \|\boldsymbol{y}\|_2,
\end{aligned}
$$

where the last inequality follows from the triangle inequality.

(b) We sketch the proof for the linear case only, the concave case being similar. Let $\boldsymbol{x}$ be a robust solution, and $\boldsymbol{v} \in \mathbb{R}^L$ so that $\delta^\star(\boldsymbol{P}^\top \boldsymbol{x}|\mathcal{Z}) = \delta^\star(\boldsymbol{v}|\mathcal{Z}_1) + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}|\mathcal{Z}_2)$.

$$
\begin{aligned}
\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) &\leq \mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > \delta^\star(\boldsymbol{P}^\top \boldsymbol{x}|\mathcal{Z}_1 \cap \mathcal{Z}_2)\right) \\
&\leq \mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{v} + \tilde{\boldsymbol{z}}^\top(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}) > \delta^\star(\boldsymbol{v}|\mathcal{Z}_1) + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}|\mathcal{Z}_2)\right) \\
&= \mathbb{P}\left(\tilde{\boldsymbol{z}}^\top(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}) > \underbrace{\delta^\star(\boldsymbol{v}|\mathcal{Z}_1) - \tilde{\boldsymbol{z}}^\top \boldsymbol{v}}_{>0 \text{ for } \tilde{\boldsymbol{z}} \in \mathcal{Z}_1} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}|\mathcal{Z}_2)\right) \\
&\leq \mathbb{P}\left(\tilde{\boldsymbol{z}}^\top(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}) > \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}|\mathcal{Z}_2)\right).
\end{aligned}
$$

Applying a Chernoff bound and uniformly bounding the right hand side concludes the proof. □

Remark 3 The proof of Proposition 7(b) is informative for a posteriori bound as well. Indeed, we have

$$
\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \exp\left(-\frac{\delta^\star(\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}|\mathcal{Z})^2}{2\|\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}\|_2^2}\right),
$$

which might be tighter than the bound from Theorem 1.

Example 4 The budget uncertainty set. The budget uncertainty set defined as

$$
\mathcal{Z}_{\ell_\infty \cap \ell_1} := \left\{\boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{z}\|_\infty \leq 1, \|\boldsymbol{z}\|_1 \leq \rho\right\},
$$

with $1 \leq \rho \leq L$ and introduced by Bertsimas and Sim [5] is the intersection of the ℓ_∞ unit-ball with an ℓ_1-ball. The robust constraint

$$
\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \boldsymbol{z}^\top \boldsymbol{P}^\top \boldsymbol{x} \leq b, \forall \boldsymbol{z} \in \mathcal{Z}_{\ell_\infty \cap \ell_1}
$$

is then equivalent to

$$
\exists \boldsymbol{v}, \; \bar{\boldsymbol{a}}^\top \boldsymbol{x} + \|\boldsymbol{v}\|_1 + \Gamma\|\boldsymbol{P}^\top \boldsymbol{x} - \boldsymbol{v}\|_\infty \leq b,
$$

requiring the introduction of the L new variables $\boldsymbol{v}$, in addition to $L+1$ auxiliary variables and $4L$ constraints to linearize the ℓ_1 and ℓ_∞ norms. With m constraints involving the same uncertain vector $\boldsymbol{a}$, there is a total of $(2L+1)m$ new variables and $4Lm$ new constraints needed. According to

Proposition 7(b), $\mathcal{Z}_{\ell_\infty \cap \ell_1}$ induces a probabilistic guarantee of $\exp\left(-\rho^2/2L\right)$, recovering the original result from Bertsimas and Sim [5]. Our result holds whenever $\tilde{\boldsymbol{z}}$ is sub-Gaussian with variance proxy 1 (Assumption 1) and $\|\tilde{\boldsymbol{z}}\|_\infty \leq 1$, but does not require $\tilde{\boldsymbol{z}}$ to be symmetrically distributed. Similarly, for the so-called box-ellipsoidal uncertainty set from Ben-Tal and Nemirovski [2]

$$\mathcal{Z}_{\ell_\infty \cap \ell_2} := \left\{ \boldsymbol{z} \in \mathbb{R}^L : \|\boldsymbol{z}\|_\infty \leq 1, \|\boldsymbol{z}\|_2 \leq \rho \right\},$$

we recover their $e^{-\rho^2/2}$ guarantee.

4.3 Minkowski sumof norm balls

We now provide explicit a priori probabilistic guarantees for sets defined as the Minkowski sum of norm balls, later referred to as sum sets. Compared to intersection sets, sum sets have not received the attention they deserve, despite their improved tractability.

Proposition 8 *Let $\mathcal{Z}_i$, for $i = 1, 2$, be two uncertainty sets. Then*

$$\rho(\mathcal{Z}_1 + \mathcal{Z}_2) \geq \rho(\mathcal{Z}_1) + \rho(\mathcal{Z}_2).$$

Proof If $\mathcal{Z} = \mathcal{Z}_1 + \mathcal{Z}_2$, then $\delta^\star(\boldsymbol{y}|\mathcal{Z}) = \delta^\star(\boldsymbol{y}|\mathcal{Z}_1) + \delta^\star(\boldsymbol{y}|\mathcal{Z}_2)$ [4, Lemma 6.3]. Hence,

$$\delta^\star(\boldsymbol{y}|\mathcal{Z}) \geq \rho(\mathcal{Z}_1)\|\boldsymbol{y}\|_2 + \rho(\mathcal{Z}_2)\|\boldsymbol{y}\|_2.$$

Taking the minimum over all $\boldsymbol{y}$ such that $\|\boldsymbol{y}\|_2 = 1$ concludes the proof. □

Example 5 Alternative to the budget uncertainty set. From a modeling perspective, the budget uncertainty $\mathcal{Z}_{\ell_\infty \cap \ell_1}$ set is geometrically very similar to the $\ell_\infty + \ell_1$ set, defined as

$$\mathcal{Z}_{\ell_\infty + \ell_1} := \left\{ \boldsymbol{z} \in \mathbb{R}^L : \boldsymbol{z} = \boldsymbol{z}_1 + \boldsymbol{z}_2, \text{ with } \|\boldsymbol{z}_1\|_\infty \leq \rho_1, \|\boldsymbol{z}_2\|_1 \leq \rho_2 \right\}.$$

In dimension $L = 2$, with proper scaling, those two sets are indeed identical as shown on Figure 2, while this is no longer the case in higher dimension (see Figure 3 for $L = 3$).

In terms of tractability, however, the sum is preferable over the intersection for it does not require additional variables $\boldsymbol{v}$. The constraint

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \boldsymbol{z}^\top \boldsymbol{P}^\top \boldsymbol{x} \leq b, \ \forall \boldsymbol{z} \in \mathcal{Z}_{\ell_\infty + \ell_1}$$

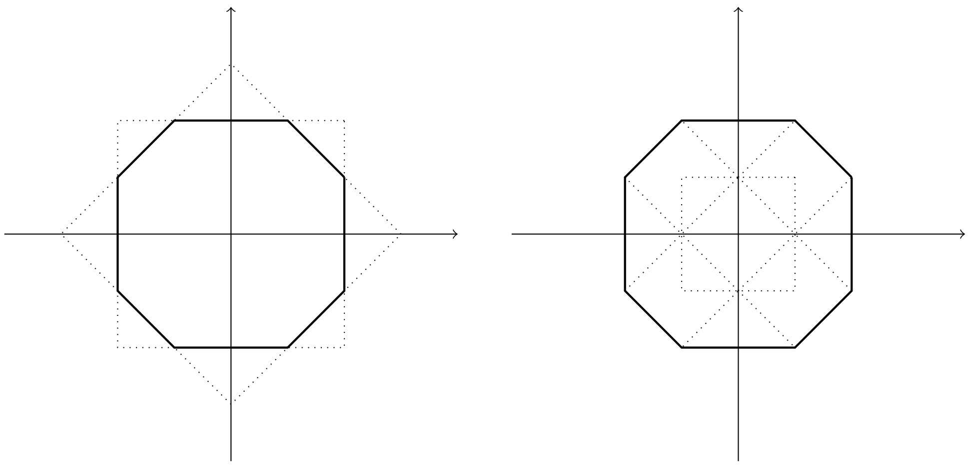
is equivalent to

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \rho_1 \|\boldsymbol{P}^\top \boldsymbol{x}\|_1 + \rho_2 \|\boldsymbol{P}^\top \boldsymbol{x}\|_\infty \leq b,$$

which requires $L + 1$ extra variables and $3L$ extra constraints to linearize the ℓ_1 and ℓ_∞ norm of $\boldsymbol{P}^\top \boldsymbol{x}$. This gain is particularly sizable when the uncertainty affects multiple constraints since the extra variables and constraints can be shared across constraints: With m uncertain constraints involving the same uncertain vector $\boldsymbol{a}$, there is a total of $(L + 1)$ new variables and $3L$ new constraints needed only. Applying Proposition 8, we can show that the $\mathcal{Z}_{\ell_\infty + \ell_1}$ set yields a probabilistic guarantee of

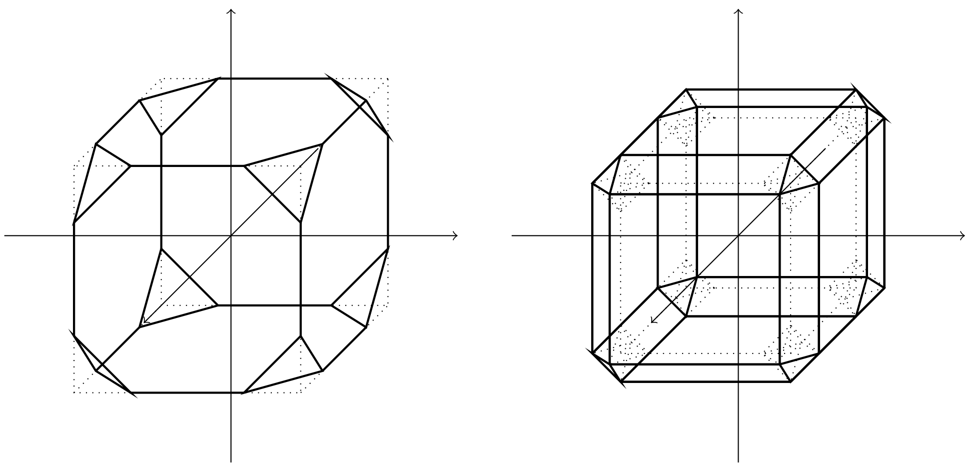
$$\exp\left(-\frac{1}{2} \left\{ \rho_1 + \rho_2 L^{-1/2} \right\}^2 \right).$$

Consequently, in the case of the ℓ_∞ and ℓ_1 norms, the sum set dominates the intersection set in terms of tractability and can still provide probabilistic guarantees. The intersection set, however, could be used for intuition and scaling of ρ_1 and ρ_2. We will concretize this comparison on a numerical example in Section 5.1.2.



(a) Intersection set with $\rho = 1.5$ (b) Sum set with $\rho_1 = \rho - 1$, $\rho_2 = 2 - \rho$.

Fig. 2: Comparison of the intersection and sum set in dimension $L = 2$.



(a) Intersection set with $\rho = 2.5$ (b) Sum set with $\rho_1 = (\rho - 1)/2$, $\rho_2 = (3 - \rho)/2$

Fig. 3: Comparison of the intersection and sum set in dimension $L = 3$.

Example 6 Alternative to the box-ellipsoidal uncertainty set. Similarly, we propose an analogous sum set for the box-ellipsoidal uncertainty set $\mathcal{Z}_{\ell_\infty \cap \ell_2}$, namely we consider

$$\mathcal{Z}_{\ell_\infty + \ell_2} := \left\{ \boldsymbol{z} \in \mathbb{R}^L : \boldsymbol{z} = \boldsymbol{z}_1 + \boldsymbol{z}_2, \text{ with } \|\boldsymbol{z}_1\|_\infty \leq \rho_1, \|\boldsymbol{z}_2\|_2 \leq \rho_2 \right\}.$$

For similar reasons, the robust counterpart associated with $\mathcal{Z}_{\ell_\infty + \ell_2}$ is more tractable in terms of additional variables and constraints. In addition, according to Proposition 8, it yields a probabilistic guarantee of

$$\exp\left(-\frac{1}{2} \left\{ \rho_1 + \rho_2 \right\}^2 \right),$$

which is independent of the number of uncertain parameters L.

Remark 4 Sets of the form $\mathcal{Z}_{\ell_\infty+\ell_2}$ have been consider by Ben-Tal et al. [3, Chapter 2] in the context of tractable reformulations of scalar chance constraints. Interestingly, sum sets emerge naturally in this context when the mean of the uncertainty quantity $\tilde{\boldsymbol{z}}$ is assumed to be bounded, rather than equal to $\mathbf{0}$ [see 3, Theorem 2.4.4 and Examples 2.4.6, 2.4.7, 2.4.9, 2.4.11]. Alternatively, uncertainty on the expected value of $\tilde{\boldsymbol{z}}$ can be modeled by decomposing $\tilde{\boldsymbol{z}}$ into $\tilde{\boldsymbol{z}}_1 + \tilde{\boldsymbol{z}}_2$, where $\tilde{\boldsymbol{z}}_1$ and $\tilde{\boldsymbol{z}}_2$ are unknown parameters with mean $\mathbf{0}$. Intuitively, given $\tilde{\boldsymbol{z}}_1$, $\tilde{\boldsymbol{z}}$ has mean $\tilde{\boldsymbol{z}}_1 \neq \mathbf{0}$ and $\tilde{\boldsymbol{z}}_2$ represents deviation from the mean. From a practical standpoint, these two interpretations lead to similar uncertainty sets.

4.4 Numerical computation using the maximum-volume inscribed sphere

For general uncertainty sets $\mathcal{Z}$, an analytical expression for the robust complexity might be out-of-reach. In this case, we propose a numerical strategy to compute valid a priori probabilistic guarantees. Indeed, we show that the robust complexity of a set is equal to the radius of the maximum-volume inscribed sphere centered at $\mathbf{0}$.

Observe that if $\mathcal{Z}' \subseteq \mathcal{Z}$ then $\mathcal{Z}$ protects against at least as many adverse scenarios as $\mathcal{Z}'$, so the probability of constraint violation with $\mathcal{Z}$ is lower than or equal to the probability of constraint violation with $\mathcal{Z}'$. In short, $\rho(\mathcal{Z}) \geq \rho(\mathcal{Z}')$. This observation suggests a numerical approach to lower bound $\rho(\mathcal{Z})$: First, compute $\mathcal{Z}'$ such that $\mathcal{Z}' \subseteq \mathcal{Z}$ and $\rho(\mathcal{Z}')$ is reasonably known. Then, use the a priori probabilistic guarantee of $\mathcal{Z}'$ for $\mathcal{Z}$. Intuitively, the closer $\mathcal{Z}'$ is to $\mathcal{Z}$, the tighter the approximation.

Actually, this procedure can produce the exact value of the robust complexity of $\mathcal{Z}$ when $\mathcal{Z}'$ is chosen as an ℓ_2-sphere centered at $\mathbf{0}$. Denote $\mathcal{B}(\mathbf{0}, \boldsymbol{r}) = \{\boldsymbol{z} : \|\boldsymbol{z}\|_2 \leq r\}$.

Proposition 9 *For any fully-dimensional convex set $\mathcal{Z} \subseteq \mathbb{R}^L$ containing $\mathbf{0}$ in its interior, the robust complexity of $\mathcal{Z}$ is the radius of the maximum inscribed sphere centered at $\mathbf{0}$ contained within $\mathcal{Z}$, i.e.,*

$$\rho(\mathcal{Z}) = \max_{r \geq 0} \quad r \text{ s.t. } \mathcal{B}(\mathbf{0}, \boldsymbol{r}) \subseteq \mathcal{Z}.$$

Proof Denote $r^\star$ the objective value of the optimization problem on the left-hand side. Note that $r^\star > 0$ since $\mathbf{0}$ lies in the interior of $\mathcal{Z}$. By definition, $\mathcal{B}(\mathbf{0}, \boldsymbol{r}^\star) \subseteq \mathcal{Z}$ so $\rho(\mathcal{Z}) \geq \rho(\mathcal{B}(\mathbf{0}, \boldsymbol{r}^\star) = r^\star$, where the last equality follows from Proposition 5. We consider a vector $\boldsymbol{z}^\star$, $\|\boldsymbol{z}^\star\|_2 = r^\star$, which lies on the boundary of $\mathcal{Z}$. Such a vector exists by optimality of $r^\star$. Then, $\boldsymbol{z}^\star$ defines a hyperplane that is tangent to $\mathcal{Z}$, hence proving that the linear optimization problem $\max_{\boldsymbol{z} \in \mathcal{Z}} \boldsymbol{z}^\top \boldsymbol{z}^\star$ admits $\boldsymbol{z} = \boldsymbol{z}^\star$ as an optimal solution. As a result, by considering $\boldsymbol{y} = \boldsymbol{z}^\star / r^\star$ we have $\rho(\mathcal{Z}) \leq r^\star$.

The optimization problem in Proposition 9 is a special case of the maximum inscribed ellipsoid [44] or the maximum inscribed sphere problem [40]. Although we fix the center of the ellipsoid to $\mathbf{0}$ and only allow for shape matrices of the form $r\boldsymbol{I}_L$, solving this optimization problem can be challenging in general. For polyhedral uncertainty sets defined with k linear constraints, $\mathcal{Z} = \{\boldsymbol{z} : \boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}\}$, however, it can be solved in closed-form:

$$\begin{aligned}
\rho(\mathcal{Z}) &= \max_{r \geq 0} r \;\text{ s.t. } r\|\boldsymbol{D}^\top \boldsymbol{e}_i\|_2 \leq d_i, \forall i = 1, \dots, k, \\
&= \min_{i=1,\dots,k} \frac{d_i}{\|\boldsymbol{D}^\top \boldsymbol{e}_i\|_2}.
\end{aligned}$$

5 Numerical experiments

In this section, we illustrate how our results materialize on two concrete examples: facility location and portfolio optimization. The first example illustrates the case where the constraints are linear. In particular, some decision variables are binary and multiple constraints are subject to uncertainty. In this context, a posteriori bound are significantly tighter due to discreteness of $\boldsymbol{x}$, hence especially valuable when considering the probability of multiple constraints being violated. We also implement and compare uncertainty regions described as intersection and sums of norm balls. The second example illustrates how our results generalize to a case where the constraint is non-linear.

5.1 Case when the constraint is linear: Facility location problem

Given a set of n potential facilities and m customers, the facility location problem consists in constructing facilities $i = 1, \dots, n$ at cost c_i in order to satisfy demand at minimal cost, i.e., solve

$$\min_{\boldsymbol{x} \in \{0,1\}^n, \boldsymbol{X} \in \mathbb{R}_+^{n \times m}} \sum_{i=1}^n c_i x_i + \sum_{j=1}^m \sum_{i=1}^n C_{ij} X_{ij} \quad \text{s.t.} \sum_{i=1}^n X_{ij} = 1, \forall j = 1, \dots, m,$$
$$\sum_{j=1}^m d_j X_{ij} \leq u_i x_i, \forall i = 1, \dots, n.$$

In this formulation, X_{ij} corresponds to the fraction of the demand of customer j produced in and shipped from facility i, at a marginal cost C_{ij}, u_i is the production capacity of facility i and d_j the demand of customer j. The first set of constraints ensures that all demand is satisfied, while the second set of constraints corresponds to production capacity constraints. The latter are linear constraints of the form "$\boldsymbol{a}^\top \boldsymbol{x} \leq b$"

$$\begin{pmatrix} -u_i \\ \boldsymbol{d} \end{pmatrix}^\top \begin{pmatrix} x_i \\ \boldsymbol{X} \boldsymbol{e}_i \end{pmatrix} \leq 0, \forall i = 1, \dots, n, \tag{6}$$

which we want to protect against uncertainty in the demand vector $\boldsymbol{d}$. Values for the nominal problem are taken from the `p1` instance of Holmberg et al. [28] with $n = 10$ facilities and $m = 50$ customers. As in Baron et al. [1], we assume that the true demand can deviate within ϵ_0 of its nominal value (we take $\epsilon_0 = 20\%$), namely for each customer $j = 1, \dots, m$,

$$\tilde{d}_j = (1 + \epsilon_0 \tilde{z}_j) \bar{d}_j,$$

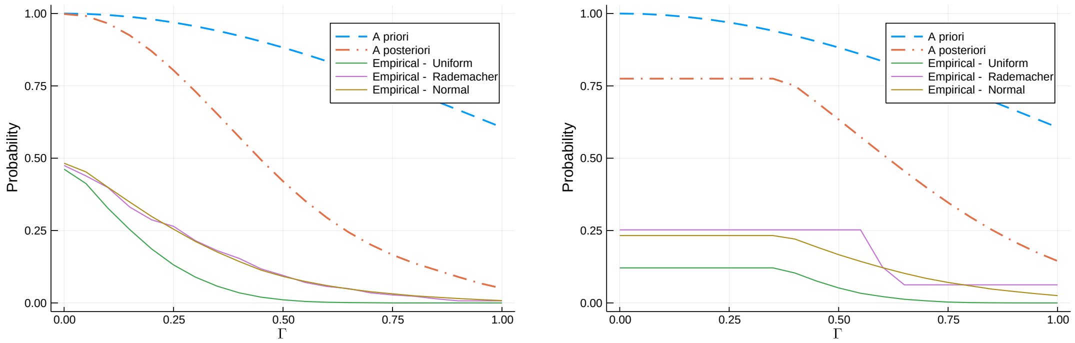
where $\tilde{\boldsymbol{z}}$ satisfies Assumption 1. For simulation purposes, we will consider three particular distributions for $\tilde{\boldsymbol{z}}$:

- **Uniform**, where each $\tilde{z}_j$ is uniformly distributed on $[-1, 1]$.
- **Normal**, where $\tilde{z}_j$'s are independently sampled from a standard distribution.
- **Rademacher**, where $\mathbb{P}(\tilde{z}_j = \pm 1) = 1/2$.

For each facility i, we replace the production capacity constraint by its robust counterpart

$$\left[\begin{pmatrix} -u_i \\ \bar{\boldsymbol{d}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{0}^\top \\ \epsilon_0 \mathrm{Diag}(\bar{\boldsymbol{d}}) \end{pmatrix} \boldsymbol{z} \right]^\top \begin{pmatrix} x_i \\ \boldsymbol{X} \boldsymbol{e}_i \end{pmatrix} \leq 0, \forall \boldsymbol{z} \in \mathcal{Z},$$

which is of the form "$[\bar{\boldsymbol{a}} + \boldsymbol{P}\boldsymbol{z}]^\top \boldsymbol{x} \leq b$", and consider different uncertainty set. In the expression above, $\mathrm{Diag}(\bar{\boldsymbol{d}})$ denotes the $m \times m$ diagonal matrix whose diagonal entries are given by $\bar{\boldsymbol{d}}$.



(a) Facility $i = 4$ (b) Facility $i = 9$

Fig. 4: Comparison of a priori and a posteriori bounds with empirical probability of constraint violation as the budget of uncertainty Γ increases, for two production capacity constraints.

5.1.1 Box uncertainty set

We first consider a box uncertainty set

$$\mathcal{Z}_\Gamma^{\text{box}} = \{ \boldsymbol{z} \in \mathbb{R}^m \ : \ \|\boldsymbol{z}\|_\infty \leq \Gamma \},$$

as in Baron et al. [1]. According to Corollary 1, the uncertainty set $\mathcal{Z}_\Gamma^{\text{box}}$ a priori induces a probabilistic guarantee of $\exp(-\rho(\mathcal{Z}_\Gamma^{\text{box}})^2/2)$ with $\rho(\mathcal{Z}_\Gamma^{\text{box}}) = \Gamma$ (Proposition 5). As for posteriori bounds, given a robust solution $(\boldsymbol{x}, \boldsymbol{X})$, we have a probabilistic guarantee of

$$\exp\left(-\frac{1}{2} \frac{|\text{Diag}(\bar{\boldsymbol{d}})^\top \boldsymbol{X} \boldsymbol{e}_i - u_i x_i|^2}{\epsilon_0^2 \|\text{Diag}(\bar{\boldsymbol{d}})^\top \boldsymbol{X} \boldsymbol{e}_i\|_2^2} \right).$$

Figure 4 compares these a priori and a posteriori bounds to the empirical probability of constraint violation for three different distributions, and two different constraints. The a posteriori bound brings a material improvement over the a priori one and better approximates the empirical probability as Γ increases. This should come as no surprise, since our bounds are consequences of concentration inequalities which are tighter as we shift further away from the mean.

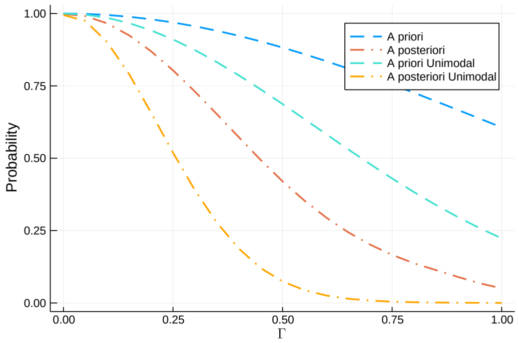
These bounds can be improved by imposing more assumptions on $\tilde{\boldsymbol{z}}$. For instance, if we assume that $\tilde{z}_j$ admits a bounded symmetric unimodal distribution (such as the uniform distribution), then, as seen in Section 3.3, $\mathcal{Z}_\Gamma^{\text{box}}$ induces an a priori guarantee of $\exp(-3\Gamma^2/2)$ instead of $\exp(-\Gamma^2/2)$, and the a posteriori guarantee is affected by a factor 3 as well. Figure 5 compares the a priori and a posteriori guarantees with and without this assumption, for the two previous constraints.

5.1.2 Budget and sum uncertainty set

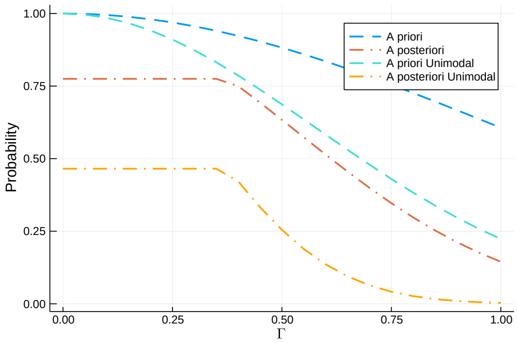
In this section, we compare the budget and sum uncertainty sets, defined as

$$\begin{aligned} \mathcal{Z}_\Gamma^{\text{budget}} &= \{ \boldsymbol{z} \in \mathbb{R}^m \ : \ \|\boldsymbol{z}\|_\infty \leq 1, \ \|\boldsymbol{z}\|_1 \leq \Gamma \}, \\ \mathcal{Z}_\Gamma^{\text{sum}} &= \{ \boldsymbol{z}_1 + \boldsymbol{z}_2 \in \mathbb{R}^m \ : \ \|\boldsymbol{z}_1\|_\infty \leq \Gamma_1, \ \|\boldsymbol{z}_2\|_1 \leq \Gamma_2 \}, \end{aligned}$$

in terms of a priori and a posteriori guarantees, as discussed in Section 4.2.

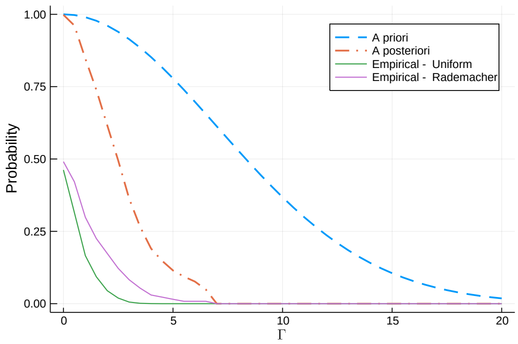


(a) Facility $i = 4$

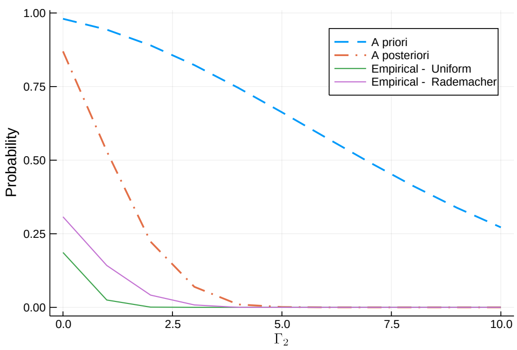


(b) Facility $i = 9$

Fig. 5: Comparison of a priori and a posteriori bounds with and without the unimodal assumption as the budget of uncertainty Γ increases, for two production capacity constraints.



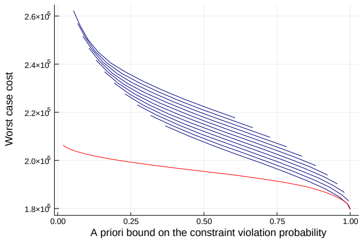
(a) Budget uncertainty set $\mathcal{Z}_\Gamma^{\text{budget}}$



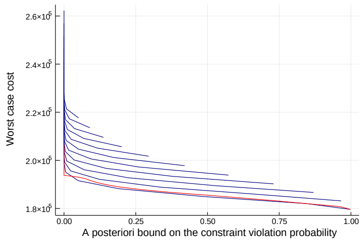
(b) Sum uncertainty set $\mathcal{Z}_\Gamma^{\text{sum}}$

Fig. 6: Comparison of a priori and a posteriori bounds with empirical probability of constraint violation for facility $i = 4$ as the budget of uncertainty Γ for the budget uncertainty set increases (left) and as Γ_2 increases for the sum uncertainty set (right), Γ_1 being fixed to 0.2.

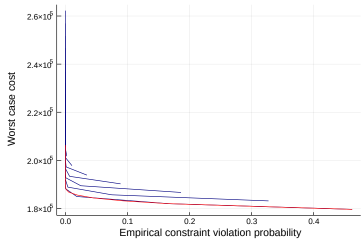
Figure 6 compares the different bounds obtained for these two uncertainty sets. Since the sum uncertainty set is parametrized by two budgets of uncertainty, we show the bounds as Γ_2 increases, Γ_1 being fixed. To derive useful conclusions, probabilistic guarantees need to be put in contrast with the conservatism of the corresponding solutions. Figure 7 represents the trade-off between the probability of constraint violation and the worst-case cost of the solution for the two uncertainty sets. For every protection level, the budget uncertainty set a priori leads (left panel) to a less conservative solution than the sum uncertainty set, by c.10%. A posteriori and empirically, this gap is generally confirmed yet weaker, in particular when Γ_1 is close to 0. This conclusion is of course valid for this particular problem only and we do not claim any generalization to other contexts. Yet, we believe that comparing uncertainty sets in terms of the trade-off between conservatism and risk level is a useful tool to inspect seemingly similar modeling choices.



(a) A priori



(b) A posteriori



(c) Empirical

Fig. 7: Comparison of the trade-off worst case cost vs. a priori (left), a posteriori probabilistic guarantee (middle) and empirical probability of constraint violation (with uniform distribution, right) for the budget (in red) and sum (in blue) uncertainty sets and varying budget of uncertainty. For the sum uncertainty sets, we connected the points corresponding to sets with the same value of Γ_1.

5.1.3 Extension to multiple constraints

For now, we only looked at the probability of one constraint being violated. To extend our analysis to multiple constraints, a simple union bound technique, alternatively called Bonferroni inequality [10], would bound the probability of at least one constraint being violated by the sum of the individual constraint violation probabilities. In the presence of k linear constraints subject to uncertainty for instance, we have

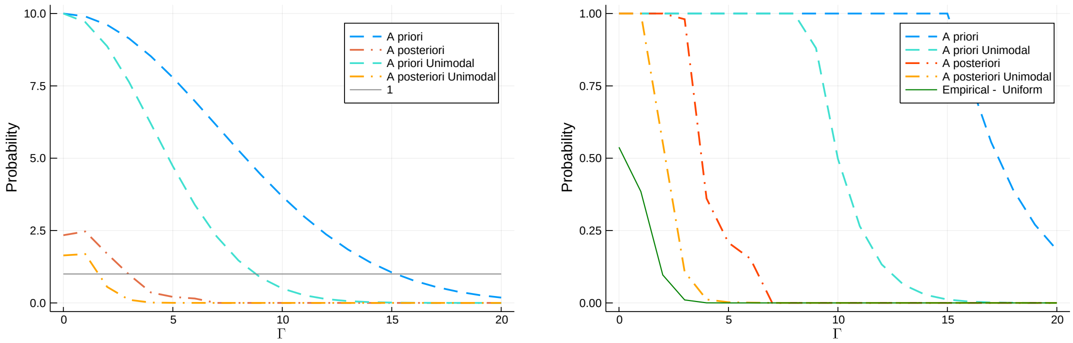
$$\mathbb{P}\left(\exists i \in \{1, \ldots, k\} \ : \boldsymbol{x}^\top \bar{\boldsymbol{a}}^{(i)} + \bar{x}^\top \boldsymbol{P}^{(i)} \tilde{\boldsymbol{z}} > 0\right) \leq \sum_{i=1}^{k} \mathbb{P}\left(\boldsymbol{x}^\top \bar{\boldsymbol{a}}^{(i)} + \bar{x}^\top \boldsymbol{P}^{(i)} \tilde{\boldsymbol{z}} > 0\right).$$

We apply this reasoning to the $k = n = 10$ production capacity constraints (6) for the budget uncertainty set $\mathcal{Z}_{\Gamma}^{\text{budget}}$. We compute a priori and a posteriori guarantees on Figure 8a, with and without the assumption that $\tilde{\boldsymbol{z}}$ is unimodal with respect to $\mathbf{0}$. For small values of Γ, these bounds, and a priori bounds especially, are clueless for they are greater than 1. A posteriori bounds on the other hand are prominently tighter. This is mainly due to the fact that many facilities are turned off ($x_i = 0$), in which case the corresponding constraints are no longer subject to uncertainty and the a posteriori bounds equal 0, whereas the a priori bounds are stricly positive. We further compare our bounds (capped at 1) with empirical probabilities (Figure 8b). These bounds are noticeably weak for small values of Γ, because the Bonferroni approximation did not account for correlations between the different constraints. Let us remark that the robust counterpart of the capacity constraints (6) similarly computes the worst case for each constraint independently.

5.2 Case when the constraint is concave in the uncertain parameter: Mean-variance portfolio

We now consider the mean-variance portfolio problem from Example 3. As in Marandi et al. [32], we use monthly average value weighted return of $n = 30$ industries from 1956 till 2015[1] to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Namely, we compute their respective maximum likelihood estimators $\bar{\boldsymbol{\mu}}$ and

[1] Available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html` under "Industry Portfolios"



(a) A posteriori vs. a priori (b) A priori vs. empirical

Fig. 8: Comparison of a priori and a posteriori bounds on the joint probability of at least one constraint being violated, with the empirical constraint violation probability, as the budget of uncertainty Γ increases.

$\bar{\boldsymbol{\Sigma}}$, and construct the uncertainty set $(\mathbb{R}^n \times \mathcal{S}_+^n) \cap \mathcal{U}$, with

$$\mathcal{U} := \left\{ (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \; : \; \begin{array}{ll} \mu_i = \bar{\mu}_i + \varepsilon_0 \bar{\mu}_i z_i, & \forall i = 1, \dots, n, \\ \Sigma_{ij} = \bar{\Sigma}_{ij} + \varepsilon_0 \bar{\Sigma}_{ij} Z_{ij}, & \forall i, j = 1, \dots, n, \\ \|\mathrm{vec}(\boldsymbol{z}, \boldsymbol{Z})\| \leq \Gamma & \end{array} \right\}.$$

In our numerical experiments, we fix $\lambda = 2$ and $\varepsilon_0 = 5\%$, vary the budget of uncertainty Γ and consider the ℓ_2, and ℓ_∞ norms to bound the magnitude of $\mathrm{vec}(z, \boldsymbol{Z})$. Note that $\mathcal{U}$ is of the desired form with $\boldsymbol{P} := \varepsilon_0 \mathrm{Diag}(\mathrm{vec}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}))$. As in the linear case, we sample $\mathrm{vec}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{Z}})$ from three distributions: Uniform within $[-1, 1]$, Rademacher and Standard distribution.

The robust constraint is of the form

$$f(\boldsymbol{a}, \boldsymbol{x}) \leq 0, \forall \boldsymbol{a} \in \mathcal{U} \text{ with } \; \boldsymbol{a} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$
$$\text{and } \; f(\boldsymbol{a}, \boldsymbol{x}) = -t - \boldsymbol{\mu}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x},$$

and the equivalent reformulation (Proposition 2) holds with

$$\boldsymbol{v} = (-\boldsymbol{x}, \boldsymbol{W}), \;\; f_\star(\boldsymbol{v}, \boldsymbol{x}) = t, \;\; \text{and } \mathcal{Z} = \{(\boldsymbol{z}, \boldsymbol{Z}) : \|\mathrm{vec}(\boldsymbol{z}, \boldsymbol{Z})\| \leq \Gamma\}.$$

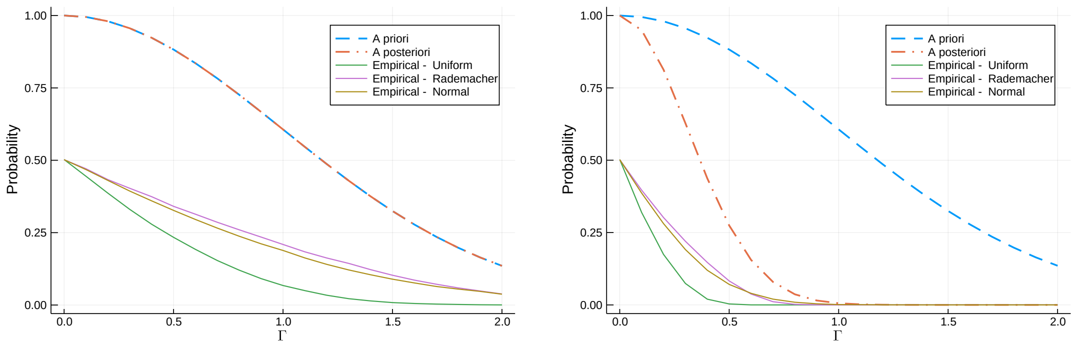
In terms of probabilistic guarantees, the a priori bound (Corollary 2) is identical to the linear case:

$$\mathbb{P}\left(t < -\tilde{\boldsymbol{\mu}}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{x}\right) \leq \exp\left(-\frac{\rho(\mathcal{Z})^2}{2}\right) = \exp\left(-\frac{\Gamma^2}{2}\right),$$

since $\rho(\mathcal{Z}) = \Gamma$ for the ℓ_2 and ℓ_∞ norm balls. As for the a posteriori bound, Theorem 2 yields

$$\mathbb{P}\left(t < -\tilde{\boldsymbol{\mu}}^\top \boldsymbol{x} + \lambda \boldsymbol{x}^\top \tilde{\boldsymbol{\Sigma}} \boldsymbol{x}\right) \leq \exp\left(-\frac{(t + \bar{\boldsymbol{\mu}}^\top \boldsymbol{x} - \langle \bar{\boldsymbol{\Sigma}}, \boldsymbol{W} \rangle)^2}{2\|\boldsymbol{P}^\top \mathrm{vec}(-\boldsymbol{x}, \boldsymbol{W})\|_2^2}\right).$$

We compare these bounds with the empirical probability of constraint violation in Figure 9. The chosen uncertainty sets, namely the ℓ_2 and ℓ_∞ balls, illustrate the two extreme sides of the spectrum: For the ℓ_2 ball (left panel), a priori and a posteriori bounds match, while for the ℓ_∞ ball (right panel), a posteriori bounds are materially tighter than a priori ones.



(a) $\mathcal{Z}$ is an ℓ_2 ball

(b) $\mathcal{Z}$ is an ℓ_∞ ball

Fig. 9: Comparison of a priori bounds, a posteriori bounds and empirical probabilities of constraint violation for the mean-variance portfolio example (Example 3) as Γ increases.

6 Concluding remarks

In this work, we developed a principled methodology for deriving strong a priori and a posteriori probabilistic guarantees for solutions of robust optimization problems. Our analysis applies broadly to any convex compact uncertainty set and to any constraint affected by uncertainty in a concave manner, and combines theoretical appeal with practical relevance.

Constraints where the uncertainty appears in a convex way are notably harder to account for in robust optimization and call for tractable safe approximations [6, 36], a topic which constitutes an exciting area for future research. In line with the present paper, probabilistic guarantees for such safe approximations would be particularly beneficial in practice. This question intimately relates to approximations of chance constrained conic or matrix inequalities [3, Chapter 10], which has received revived interest recently under the lens of distributionally robust optimization [15, 16, 45].

References

1. Opher Baron, Joseph Milner, and Hussein Naseraldin. Facility location: A robust optimization approach. *Production and Operations Management*, 20(5):772–785, 2011.
2. Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.
3. Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*, volume 28. Princeton University Press, 2009.
4. Aharon Ben-Tal, Dick den Hertog, and Jean-Philippe Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299, 2015.
5. Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1): 35–53, 2004.
6. Dimitris Bertsimas and Melvyn Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1-2):5–36, 2006.
7. Dimitris Bertsimas, Dessislava Pachamanova, and Melvyn Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.
8. Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

9. Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. Data-driven robust optimization. *Mathematical Programming*, 167(2):235–292, 2018.
10. Carlo Emilio Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62, 1936.
11. Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
12. Giuseppe Calafiore and Marco C Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.
13. Giuseppe C Calafiore and Marco C Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
14. Marco C Campi and Simone Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.
15. Wenqing Chen, Melvyn Sim, Jie Sun, and Chung-Piaw Teo. From cvar to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58(2):470–485, 2010.
16. Zhi Chen, Daniel Kuhn, and Wolfram Wiesemann. Data-driven chance constrained programs over Wasserstein balls. *arXiv preprint arXiv:1809.00210*, 2018.
17. Daniela Pucci De Farias and Benjamin Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of operations research*, 29(3):462–478, 2004.
18. Erick Delage and Shie Mannor. Percentile optimization for markov decision processes with parameter uncertainty. *Operations research*, 58(1):203–213, 2010.
19. Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.
20. Emre Erdoğan and Garud Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1-2):37–61, 2006.
21. Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.
22. Virginie Gabrel, Cécile Murat, and Aurélie Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483, 2014.
23. Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.
24. Yannis A Guzman, Logan R Matthews, and Christodoulos A Floudas. New a priori and a posteriori probabilistic bounds for robust counterpart optimization: I. unknown probability distributions. *Computers & Chemical Engineering*, 84:568–598, 2016.
25. Yannis A Guzman, Logan R Matthews, and Christodoulos A Floudas. New a priori and a posteriori probabilistic bounds for robust counterpart optimization: Ii. a priori bounds for known symmetric and asymmetric probability distributions. *Computers & Chemical Engineering*, 101:279–311, 2017.
26. Yannis A Guzman, Logan R Matthews, and Christodoulos A Floudas. New a priori and a posteriori probabilistic bounds for robust counterpart optimization: Iii. exact and near-exact a posteriori expressions for known probability distributions. *Computers & Chemical Engineering*, 103:116–143, 2017.
27. Grani A Hanasusanto, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62, 2015.
28. Kaj Holmberg, Mikael Rönnqvist, and Di Yuan. An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research*,

113(3):544–559, 1999.
29. Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pages 130–166. INFORMS, 2019.
30. Zukui Li, Qiuhua Tang, and Christodoulos A Floudas. A comparative theoretical and computational study on robust counterpart optimization: ii. probabilistic guarantees on constraint satisfaction. *Industrial & Engineering Chemistry Research*, 51(19):6769–6788, 2012.
31. James Luedtke and Shabbir Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.
32. Ahmadreza Marandi, Aharon Ben-Tal, Dick den Hertog, and Bertrand Melenberg. Extending the scope of robust quadratic optimization. *Available on Optimization Online*, 2017.
33. Arkadi Nemirovski and Alexander Shapiro. Scenario approximations of chance constraints. In *Probabilistic and randomized methods for design under uncertainty*, pages 3–47. Springer, 2006.
34. Phillippe Rigollet and Jan-Christian Hütter. High dimensional statistics. *Lecture notes for course 18.S997*, 2015. URL `http://www-math.mit.edu/~rigollet/PDFs/RigNotes17.pdf`.
35. Ralph Tyrell Rockafellar. *Convex analysis*. Princeton University Press, 2015.
36. Ernst Roos, Dick den Hertog, Aharon Ben-Tal, Frans de Ruiter, and Jianzhe Zhen. Approximation of hard uncertain convex inequalities. *Available on Optimization Online*.
37. Bart PG Van Parys, Peyman Mohajerin Esfahani, and Daniel Kuhn. From data to decisions: Distributionally robust optimization is optimal. *arXiv preprint arXiv:1704.04118*, 2017.
38. Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.
39. Weijun Xie. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming*, pages 1–41, 2019.
40. Yulai Xie, Jack Snoeyink, and Jinhui Xu. Efficient algorithm for approximating maximum inscribed sphere in high dimensional polytope. In *Proceedings of the twenty-second annual symposium on Computational Geometry*, pages 21–29, 2006.
41. Huan Xu, Constantine Caramanis, and Shie Mannor. Optimization under probabilistic envelope constraints. *Operations Research*, 60(3):682–699, 2012.
42. Wenzhuo Yang and Huan Xu. Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming*, 155(1-2):231–265, 2016.
43. İhsan Yanıkoğlu and Dick den Hertog. Safe approximations of ambiguous chance constraints using historical data. *INFORMS Journal on Computing*, 25(4):666–681, 2013.
44. Jianzhe Zhen and Dick Den Hertog. Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS Journal on Computing*, 30(1):31–42, 2018.
45. Steve Zymler, Daniel Kuhn, and Berç Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.

A Preliminary results from probability theory

In this section, we recall some useful definitions and results from probability theory regarding sub-Gaussian variables. We refer to Rigollet and Hütter [34] and Wainwright [38] for a comprehensive treatment.

Definition 3 [Definition 1.2 in 34] A random variable $\tilde{x} \in \mathbb{R}$ is said to be *sub-Gaussian* with variance proxy σ^2, denoted $\tilde{x} \sim \text{subG}\left(\sigma^2\right)$, if $\mathbb{E}[\tilde{x}] = 0$ and for all $s \in \mathbb{R}$,

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

Naturally, centered Gaussian random variables are also sub-Gaussian. Of particular interest for the RO literature, bounded random variables are a special case of sub-Gaussian random variables, a consequence of Hoeffding's inequality.

Lemma 1 *[Lemma 1.8 in 34] Let $\tilde{x}$ be a random variable such that $\mathbb{E}[\tilde{x}] = 0$ and $\tilde{x} \in [a, b]$ almost surely. Then, $\tilde{x} \sim subG\left(\frac{(b-a)^2}{4}\right)$.*

Note that the definition of sub-Gaussian random variables is essentially a Gaussian bound on the moment generating function of $\tilde{x}$ and is very similar to Property P2 in Ben-Tal et al. [3], Section 2.4. Examples from Table 2.3. in Ben-Tal et al. [3] for which $\mu_\pm = 0$ satisfy Definition 3. This definition can be extended to non-centered random variables as well [see 38, Definition 2.2] and account for examples with $\mu_\pm \neq 0$.

Bound on the moment generating function leads to bound on the tail of the distribution.

Proposition 10 *[Lemma 1.3 in 34] If $\tilde{x} \sim subG(\sigma^2)$, for all $t > 0$,*

$$\mathbb{P}(\tilde{x} > t) \leq e^{-\frac{t^2}{2\sigma^2}},$$
$$\mathbb{P}(\tilde{x} < -t) \leq e^{-\frac{t^2}{2\sigma^2}}.$$

Actually, all these tail bounds can be used as equivalent definitions of sub-Gaussian random variables [38, Theorem 2.6].

Finally, relevant for our analysis is the fact that linear combinations of independent sub-Gaussian random variables are themselves sub-Gaussian.

Proposition 11 *Let $\tilde{x}_i$, $i = 1, \ldots, n$, be n independent random variables and $\theta \in \mathbb{R}^n$.*
(a) If $\tilde{x}_i \sim subG(\sigma^2)$, then $\sum_i \theta_i \tilde{x}_i \sim subG(\|\theta\|_2^2 \sigma^2)$.
(b) If $E[\tilde{x}_i] = 0$ and $|\tilde{x}_i| \leq 1$ almost surely, then $\sum_i \theta_i \tilde{x}_i \sim subG(\|\theta\|_2^2)$.

Proof (a) The first part of the theorem follows from the definition of sub-Gaussian random variables and a Chernoff bound as proved in [34, Theorem 1.6]. (b) For the second part, Lemma 1 states that $\tilde{x}_i \sim \text{subG}(1)$. Applying Theorem 1(a) yields $\sum_i \theta_i \tilde{x}_i \sim \text{subG}(\|\theta\|_2^2)$. □

B Polynomial probabilistic guarantees

In this section, we provide polynomial probabilistic guarantees under a less restrictive set of assumptions:

Assumption 2 *We assume that the coordinates of the uncertain parameter $\tilde{\boldsymbol{z}} \in \mathbb{R}^L$ are L random variables with zero mean and covariance matrix bounded by some semi-definite positive matrix $\boldsymbol{\Sigma} \succeq 0$, i.e., $\mathbb{E}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^\top] \preceq \boldsymbol{\Sigma}$.*

We consider the case where the constraint is linear, $f(\boldsymbol{a}, \boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} - b$, and treat sequentially the a posteriori and a priori bounds. We first derive a posteriori guarantees that depend on the robust solution $\boldsymbol{x}$.

Theorem 3 *Under Assumption 2, for any $\boldsymbol{x} \in \mathcal{X}$ satisfying*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z}) \leq b,$$

we have

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \frac{\|\boldsymbol{\Sigma}^{1/2} \boldsymbol{P}^\top \boldsymbol{x}\|_2^2}{(b - \bar{\boldsymbol{a}}^\top \boldsymbol{x})^2} \leq \frac{\|\boldsymbol{\Sigma}^{1/2} \boldsymbol{P}^\top \boldsymbol{x}\|_2^2}{\delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z})^2}.$$

Proof We follow the same proof as for Theorem 1. Instead of a Chernoff bound, we apply a Chebyshev's inequality

$$\mathbb{P}\left(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > t\right) \leq \frac{\text{Var}(\tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x})}{t^2} \leq \frac{\|\boldsymbol{\Sigma}^{1/2} \boldsymbol{P}^\top \boldsymbol{x}\|_2^2}{t^2},$$

with $t = b - \bar{\boldsymbol{a}}^\top \boldsymbol{x} \geq \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z})$.

Denote $\lambda_{\max}(\boldsymbol{\Sigma})$ the maximum eigenvalue value of $\boldsymbol{\Sigma}$. Then, for any vector $\boldsymbol{y}$, $\|\boldsymbol{\Sigma}^{1/2} \boldsymbol{y}\|_2^2 \leq \lambda_{\max}(\boldsymbol{\Sigma}) \|\boldsymbol{y}\|_2^2$, and we can derive a priori probabilistic guarantees.

Proposition 12 *Under Assumption 2, for any $\boldsymbol{x} \in \mathcal{X}$ satisfying*

$$\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \delta^\star(\boldsymbol{P}^\top \boldsymbol{x} | \mathcal{Z}) \leq b,$$

we have

$$\mathbb{P}\left(\bar{\boldsymbol{a}}^\top \boldsymbol{x} + \tilde{\boldsymbol{z}}^\top \boldsymbol{P}^\top \boldsymbol{x} > b\right) \leq \lambda_{\max}(\boldsymbol{\Sigma})\, \rho(\mathcal{Z})^{-2}.$$

We omit the proof for concision. Notice that the a priori bound is again driven by the robust complexity of the uncertainty set, $\rho(\mathcal{Z})$, yet in a polynomial way. The proof techniques can be straightforwardly adapted to the general case where $\boldsymbol{f}(\boldsymbol{a}, \boldsymbol{x})$ is a concave function of $\boldsymbol{a}$.