MIT Open Access Articles

Probabilistic Guarantees in Robust Optimization

PROBABILISTIC GUARANTEES IN ROBUST OPTIMIZATION[*]

DIMITRIS BERTSIMAS[†], DICK DEN HERTOG[‡], AND JEAN PAUPHILET[§]

Abstract. We develop a general methodology for deriving 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, under minimal assumptions on the underlying stochastic process. Namely, we assume that the coordinates of the noise vector are light-tailed (sub-Gaussian) but not necessarily independent. We introduce the notion of robust complexity of an uncertainty set, which is a robust analogue of the Rademacher and Gaussian complexities encountered in high-dimensional statistics, and which connects the geometry of the uncertainty set with an a priori probabilistic guarantee. Interestingly, the robust complexity involves the support function of the uncertainty set, which also plays a crucial role in the robust counterpart theory for robust linear and nonlinear optimization. For a variety of uncertainty sets of practical interest, we are able to compute it in closed form or derive valid approximations. Our methodology recovers most of the results available in the related literature using first principles and extends them to new uncertainty sets and nonlinear constraints. We also derive improved a posteriori bounds, i.e., significantly tighter bounds which depend on the resulting robust solution.

Key words. robust optimization, support function, uncertainty set, concentration inequality

AMS subject classifications. 90C25, 65K05

DOI. 10.1137/21M1390967

1. Introduction. Over the past decades, robust optimization (RO) has emerged as an effective framework accounting for uncertainty in optimization problems (see [5, 22] for a review). Consider a single linear constraint of the form

$$(1.1) \qquad \mathbf{a}^\top \mathbf{x} \le b,$$

where the parameter $\mathbf{a}$ is subject to uncertainty, and $b \in \mathbb{R}$ is certain, without loss of generality. In a robust approach, $\mathbf{a}$ is described as a deterministic yet unknown vector belonging to a so-called uncertainty set $\mathcal{U}$, and constraint (1.1) is imposed to hold for all values of $\mathbf{a} \in \mathcal{U}$, i.e., we consider its robust counterpart

$$(1.2) \qquad \mathbf{a}^\top \mathbf{x} \le b \ \forall \mathbf{a} \in \mathcal{U}.$$

For theoretical as well as practical considerations, a central question in the RO literature has been how to derive probabilistic guarantees for robust solutions, namely how to bound the probability of constraint violation, $\mathbb{P}_{\tilde{\mathbf{a}}} \left(\tilde{\mathbf{a}}^\top \mathbf{x} > b \right)$, for $\mathbf{x}$ satisfying (1.2). Historically, connecting a deterministic model of uncertainty (1.2) with probability theory was needed to endorse RO as a competitive framework compared to stochastic optimization.

In practice, a priori bounds, i.e., bounds that only depend on the uncertainty set $\mathcal{U}$ without knowledge of the robust solution $\mathbf{x}$, are useful for calibrating and sizing the uncertainty set according to a desired protection level. Clearly, $\mathbb{P}_{\tilde{\mathbf{a}}} \left(\tilde{\mathbf{a}}^\top \mathbf{x} > b \right) \le$

[†]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (dbertsim@mit.edu).
[‡]Faculty of Economics and Business, University of Amsterdam, Amsterdam, 1001 NL, Netherlands (d.denhertog@uva.nl).
[§]London Business School, London, NW1 4SA, UK (jpauphilet@london.edu).

$\mathbb{P}_{\tilde{\mathbf{a}}}\,(\tilde{\mathbf{a}} \notin \mathcal{U})$, yet much tighter bounds can be obtained. For instance, the authors of [4] proved that $\mathbb{P}_{\tilde{\mathbf{a}}}\,(\tilde{\mathbf{a}}^\top \mathbf{x} > b) \leq \exp(-\rho^2/2)$ for their proposed box-ellipsoidal uncertainty set $\mathcal{U} = \{\mathbf{a} : \|\mathbf{a}\|_\infty \leq 1,\ \|\mathbf{a}\|_2 \leq \rho\}$, under the assumption that the coordinates of $\tilde{\mathbf{a}}$ 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 $\mathbb{P}_{\tilde{\mathbf{a}}}\,(\tilde{\mathbf{a}} \notin \mathcal{U})$. To guarantee that the robust solution is feasible with probability at least $1 - \varepsilon$, for example, this bound prescribes taking $\rho = \sqrt{-2\ln(\varepsilon)}$.

A posteriori bounds, i.e., bounds that depend on the robust solution $\mathbf{x}$, are also useful for auditing the resulting solution after it has been computed and assessing its performance.

Results for other classes of uncertainty sets $\mathcal{U}$, or a more general constraint, such as $f(\mathbf{a}, \mathbf{x}) \leq 0$, have been derived, yet in a disparate and unsystematic way. Attempts have been made to report existing bounds in a unified manner [35, 26, 27, 28] for various uncertainty sets. 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, has provided a disciplined methodology that enables us 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 guarantee is still needed to understand the connection between robust optimization and its out-of-sample performance on random instances. Such is the contribution of the present paper.

1.1. Literature review. Safe approximation of the ambiguous chance constraint. In this setting, the goal is to reformulate a given chance constraint $\mathbb{P}_{\tilde{\mathbf{a}}}\,(\tilde{\mathbf{a}}^\top \mathbf{x} \leq b) \geq 1 - \varepsilon$, with $\varepsilon > 0$. In most cases, the distribution of $\tilde{\mathbf{a}}$ is not known precisely but rather assumed to belong to a certain class, and the objective is to reformulate the ambiguous chance constraint

$$(1.3) \qquad\qquad \inf_{\mathbb{P}_{\tilde{\mathbf{a}}} \in \mathcal{P}} \mathbb{P}_{\tilde{\mathbf{a}}}\,\left(\tilde{\mathbf{a}}^\top \mathbf{x} \leq b\right) \geq 1 - \varepsilon,$$

where $\mathcal{P}$ is a class of allowable probability distributions for $\tilde{\mathbf{a}}$. Given some assumption on $\mathcal{P}$, the ambiguous chance constraint (1.3) can be proven to be equivalent to a robust constraint of the form (1.2), where the uncertainty set $\mathcal{U}$ depends on $\mathcal{P}$. Notably, exact reformulations have been derived in cases where $\tilde{\mathbf{a}}$ is normally distributed [17, 47, 48] or has known mean and bounded second-order moments [47, 48, 30]. The authors of [3, Chapter 2] derive such reformulations for a general class of probability distributions and in a disciplined manner. For example, if the coordinates of $\tilde{\mathbf{a}}$ 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 (1.2) with

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

yields $\mathbb{P}_{\tilde{\mathbf{a}}}\,(\tilde{\mathbf{a}}^\top \mathbf{x} > b) \leq \varepsilon$ [3, Theorem 2.4.4 applied to Example 2.4.7]. This approach, which takes assumptions on the distribution of $\tilde{\mathbf{a}}$ as the primitive and provides a corresponding uncertainty set, has two main shortcomings. First, it cannot provide any probabilistic guarantee for the uncertainty set $\mathcal{U}$ whenever $\tilde{\mathbf{a}}$ satisfies different assumptions. Instead, different assumptions on $\tilde{\mathbf{a}}$ 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, their approach systematically results in uncertainty sets involving the Euclidean norm, which

is closely related to the notion of variance but excludes more general uncertainty sets, such as polyhedral or budget uncertainty sets [8], despite their tractability and wide use in practice. More recently, the authors of [6] propose a data-driven extension of this procedure. Starting from assumptions on the distribution of $\tilde{\mathbf{a}}$ and given past observations $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(N)}$, 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. The latter limitation is partially addressed by [32], which develops a data-driven approach for calibrating a set $\mathcal{U}_\varepsilon$ such that $\mathbb{P}(\tilde{\mathbf{a}} \notin \mathcal{U}_\varepsilon) \leq \varepsilon$, leading to an ε-level probabilistic guarantee for any robust constraint. However, as mentioned in the introduction, the resulting uncertainty set can be prohibitively large.

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 (1.3) have been derived in cases where the ambiguity set $\mathcal{P}$ is constructed from bounds on moments of the distribution [18, 24, 51, 48] or is defined as a ball around the empirical distribution according to the f-divergence [49, 38], the Wasserstein distance [23, 15, 21, 45, 34], or the relative entropy [43]. We refer the reader to [29] and references therein for a comprehensive review. Despite powerful out-of-sample performance guarantees and moderate conservatism, DRO approaches are usually more computationally expensive than simple robust approaches, especially in data-driven settings. Notorious exceptions to this claim are [38], which solved DRO problems using first-order methods, and [21], which formulated DRO problems with the Wasserstein distance as linear and second-order cone problems. Finally, Gupta [25] extends the data-driven approach of [6] to ambiguity sets; specifically, he considers a DRO setting with a parametric class of distributions and uses Bayesian updating of the parameters of the distribution to construct relevant ambiguity sets. However, this approach suffers from limitations similarly to [6].

Scenario approach. A general way to deal with chance constraints is the scenario approach, i.e., generate N samples $\mathbf{a}^{(i)}$, $i = 1, \ldots, N$, of $\tilde{\mathbf{a}}$ and replace the chance constraint by N deterministic constraints $f(\mathbf{a}^{(i)}, \mathbf{x}) \leq 0$. Assuming that the function f is convex in $\mathbf{x}$, [11, 12] prove that this approach yields a feasible solution to the chance constrained problem with probability $1 - \delta$ (over the N samples) if

$$N \geq 2L + \frac{2L}{\varepsilon} \ln\left(\frac{2}{\varepsilon}\right) + \frac{2}{\varepsilon} \ln\left(\frac{1}{\delta}\right).$$

This approach was later refined [39, 36, 13, 33] and extended to ambiguous chance constraints [20, 16]. The scenario approach, however, suffers from two limitations. First, while the number of constraints to be sampled, N, scales favorably with respect to the protection level ε, it grows linearly in the dimension of the uncertainty, L, which makes the approach difficult to apply to medium-sized problems. Second, the probabilistic guarantee is obtained with respect to the sampling distribution, which should be as close as possible to the "true" and potentially unknown distribution of $\tilde{\mathbf{a}}$.

1.2. Contributions and structure. In this paper, we start from a general robust constraint,

$$f(\mathbf{a}, \mathbf{x}) \leq 0 \,\forall \mathbf{a} \in \mathcal{U}, \quad \Longleftrightarrow \quad \max_{\mathbf{a} \in \mathcal{U}} f(\mathbf{a}, \mathbf{x}) \leq 0,$$

and derive valid probabilistic guarantees for any solution $\mathbf{x}$ satisfying the robust constraint. We will assume that $\mathcal{U}$ is convex and $f(\mathbf{a}, \mathbf{x})$ is concave in $\mathbf{a}$ for any $\mathbf{x}$. Under this assumption, the maximization problem above is well defined, and the robust constraint can be reformulated in a tractable way [2].[1] We restrict our attention to uncertainty sets $\mathcal{U}$ of the form

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

where $\bar{\mathbf{a}}$ is the nominal value of $\mathbf{a}$, and $\mathcal{Z}$ is a given nonempty, fully dimensional convex and compact set, with $\mathbf{0} \in \mathrm{ri}(\mathcal{Z})$, as in [2]. Our main contribution is a simple methodology for deriving probabilistic guarantees for solutions of robust optimization problems, which provides a unifying perspective on existing results from the literature and extends them significantly.

A unified perspective on the literature. Our approach unifies existing results from the literature in terms of assumptions, proof techniques, and result statements. First, we observe that most results from the literature assume that the coordinates of $\tilde{\mathbf{z}}$ are sub-Gaussian and independent random variables. Under these generic assumptions, probabilistic guarantees follow directly from elementary concentration inequalities. In particular, for linear constraints of the form $\mathbf{a}(\mathbf{z})^\top \mathbf{x} \leq b(\mathbf{x}) \,\forall \mathbf{z} \in \mathcal{Z}$, we show that an uncertainty set $\mathcal{Z}$ yields a probabilistic guarantee of

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

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

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

This single and simple result recovers all known probabilistic guarantees for robust linear constraints, with minimal assumptions on the underlying distribution of the uncertain parameter. In particular, for the box-ellipsoidal set of [4] and the budget uncertainty set of [8], we show that $\rho(\mathcal{Z}) \geq \Gamma$ and $\rho(\mathcal{Z}) \geq \Gamma/\sqrt{L}$, respectively, and hence show matching existing bounds.

Significant extensions. As summarized in Table 1, our approach is also a powerful tool for extending existing results in significant ways:

- First, we readily generalize all results to the case where the constraint is nonlinear in the uncertainty. Under the same assumptions, for constraints of the form $f(\mathbf{a}, \mathbf{x}) \leq 0$, where f is concave in $\mathbf{a}$, we show that an uncertainty set $\mathcal{Z}$ yields the same probabilistic guarantee of

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

---

[1]When $f(\mathbf{a}, \mathbf{x})$ is convex in $\mathbf{a}$, however, an equivalent tractable reformulation of the robust constraint is out of reach. Safe approximations based on scenario sampling [12] or linear approximation [9] have been proposed but remain significantly less tractable. Correspondingly, the probabilistic guarantees obtained in the convex case would require more stringent assumptions on the underlying distribution; see, e.g., Conjecture 10.1 in [3].

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 $\delta^{\star}(\mathbf{y}|\mathcal{Z}) = \max_{\mathbf{z}\in\mathcal{Z}} \mathbf{z}^{\top}\mathbf{y}$ and $\rho(\mathcal{Z}) = \min_{\mathbf{y}:\|\mathbf{y}\|_2=1} \delta^{\star}(\mathbf{y}|\mathcal{Z})$. For the concave case, $f_{\star}(\mathbf{v},\mathbf{x}) := \inf_{\mathbf{a}}\left[\mathbf{a}^{\top}\mathbf{v} - f(\mathbf{a},\mathbf{x})\right]$ is often referred to as the concave conjugate of $f(\cdot,\mathbf{x})$.

Constraint type	Linear	Concave in $\mathbf{a}$
Nominal constraint	$\mathbf{a}^{\top}\mathbf{x} \leq b(\mathbf{x})$	$f(\mathbf{a},\mathbf{x}) \leq 0$
Robust counterpart	$\bar{\mathbf{a}}^{\top}\mathbf{x} + \delta^{\star}(\mathbf{P}^{\top}\mathbf{x}|\mathcal{Z}) \leq b$	$\exists\mathbf{v},\ \bar{\mathbf{a}}^{\top}\mathbf{v} + \delta^{\star}(\mathbf{P}^{\top}\mathbf{v}|\mathcal{Z}) - f_{\star}(\mathbf{v},\mathbf{x}) \leq 0$
A priori probabilistic guarantee (Corollary 3.9, 3.11)	$\exp\left(-\frac{1}{2}\rho(\mathcal{Z})^2\right)$	$\exp\left(-\frac{1}{2}\rho(\mathcal{Z})^2\right)$
A posteriori probabilistic guarantee (Theorem 3.4, 3.10)	$\exp\left(-\frac{(b(\mathbf{x})-\bar{\mathbf{a}}^{\top}\mathbf{x})^2}{2\|\mathbf{P}^{\top}\mathbf{x}\|_2^2}\right)$	$\exp\left(-\frac{(f_{\star}(\mathbf{v},\mathbf{x})-\bar{\mathbf{a}}^{\top}\mathbf{v})^2}{2\|\mathbf{P}^{\top}\mathbf{v}\|_2^2}\right)$

- We show how to tighten these **x**-independent probabilistic guarantees a posteriori, i.e., given the resulting robust solution **x**. In particular, in the linear case, we prove an a posteriori bound of the form

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

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 be computed for any general convex uncertainty set and serve as a useful auditing tool for the robustness of a solution.
- As summarized in Table 2, we are able to compute in closed form or derive valid approximations of the robust complexity $\rho(\mathcal{Z})$ for a variety of uncertainty sets of practical interest. In particular, we provide the first results for polyhedra defined with a finite number of linear inequalities and for 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.
- Finally, we analyze the case where the coordinates of $\tilde{\mathbf{z}}$ are no longer assumed to be independent. For sub-Gaussian uncertainty, we derive exponential probabilistic guarantees that still depend on the robust complexity, yet at a weaker rate. Among others, we prove that

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

We also derive polynomial bounds when only assumptions on the first two moments of $\tilde{\mathbf{z}}$ are made.

The rest of the paper is structured as follows. Section 2 recalls equivalent reformulations for general uncertain constraints. When $\tilde{\mathbf{z}}$ has independent sub-Gaussian coordinates, we provide a posteriori and a priori bounds for general constraints and uncertainty sets in section 3. Section 4 provides some closed form expressions and valid approximations for the robust complexity of a set, which drives a priori guarantees. In section 5, we challenge the independence assumption on the coordinates of $\tilde{\mathbf{z}}$ and provide two sets of generalized results: when the components of $\tilde{\mathbf{z}}$ are not necessarily independent and when $\tilde{\mathbf{z}}$ has zero mean and bounded second moments. Finally, we illustrate how our results can be applied in practice on a facility location problem in section 6.

TABLE 2

Valid lower bound on robust complexity of $\mathcal{Z}$, defined as $\rho(\mathcal{Z}) := \min_{\mathbf{y}:\|\mathbf{y}\|_2=1} \max_{\mathbf{z}\in\mathcal{Z}} \mathbf{y}^\top \mathbf{z}$, for some common uncertainty sets. For each set, we also provide sufficient conditions (s.c.) to obtain a probabilistic guarantee of at least ε, i.e., $\rho(\mathcal{Z}) \geq \gamma(\varepsilon) := \sqrt{2\ln(1/\varepsilon)}$. Instances denoted by * are valid under the extra assumption that the true uncertain parameter $\tilde{\mathbf{z}}$ satisfies $\|\tilde{\mathbf{z}}\|_\infty \leq 1$.

Uncertainty set	Definition		$\rho(\mathcal{Z}) \geq \ldots$	s.c. for ε-guarantee	Ref.
Norm-set	$\{\mathbf{z} : \|\mathbf{z}\|_p \leq \Gamma\}$	if $p \geq 2$: if $p \leq 2$:	Γ $\Gamma L^{1/2-1/p}$	$\Gamma \geq \gamma(\varepsilon)$ $\Gamma \geq L^{1/p-1/2}\gamma(\varepsilon)$	Sec. 4.1
Budget set*	$\{\mathbf{z} : \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_1 \leq \Gamma\}$		$\Gamma/\sqrt{L}$	$\Gamma \geq \sqrt{L}\gamma(\varepsilon)$	[8]
Box-ellipsoidal set*	$\{\mathbf{z} : \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_2 \leq \Gamma\}$		Γ	$\Gamma \geq \gamma(\varepsilon)$	[3]
$\ell_\infty + \ell_1$ set	$\{\mathbf{z}_1 + \mathbf{z}_2 : \|\mathbf{z}_1\|_\infty$ $\leq \rho_1, \|\mathbf{z}_2\|_1 \leq \rho_2\}$		$\rho_1 + \rho_2/\sqrt{L}$	$\rho_1 \geq \alpha\gamma(\varepsilon),\ \alpha \in [0,1]$ $\rho_2 \geq (1-\alpha)\sqrt{L}\gamma(\varepsilon)$	Sec. 4.3
$\ell_\infty + \ell_2$ set	$\{\mathbf{z}_1 + \mathbf{z}_2 : \|\mathbf{z}_1\|_\infty$ $\leq \rho_1, \|\mathbf{z}_2\|_2 \leq \rho_2\}$		$\rho_1 + \rho_2$	$\rho_1 \geq \alpha\gamma(\varepsilon),\ \alpha \in [0,1]$ $\rho_2 \geq (1-\alpha)\gamma(\varepsilon)$	[3]
Polyhedral set	$\{\mathbf{z} : \mathbf{Dz} \leq \mathbf{d}\}$		$\min_i \frac{d_i}{\|\mathbf{D}^\top \mathbf{e}_i\|_2}$	$d_i \geq \|\mathbf{D}^\top \mathbf{e}_i\|_2\gamma(\varepsilon)\ \forall i$	Sec. 4.4

Notation. In the remainder of the paper, we use nonbold (x), lowercase bold ($\mathbf{x}$), and uppercase bold ($\mathbf{X}$) 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 $\mathbf{e}_i$ denote 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 $\mathbf{x} \in \mathbb{R}^n$ as $\|\mathbf{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$. For $p = \infty$, $\|\mathbf{x}\|_\infty := \max_i |x_i|$. Any function $f : \mathbb{R}^n \to \mathbb{R}$ is defined on a subset of $\mathbb{R}^n$, called its domain and denoted $\mathrm{dom}(f)$. Similarly, f can be extended to the whole space $\mathbb{R}^n$ by setting $f(\mathbf{x}) = \infty$ if $\mathbf{x} \notin \mathrm{dom}(f)$. For any convex nonempty set $\mathcal{Z}$, its relative interior is defined and denoted by $\mathrm{ri}(\mathcal{Z}) := \{\mathbf{x} \in \mathcal{Z} : \forall \mathbf{y} \in \mathcal{Z}\ \exists \lambda > 1, \mathbf{y} + \lambda(\mathbf{x} - \mathbf{y}) \in \mathcal{Z}\}$. We denote by $\mathcal{S}_+^n$ the cone of $n \times n$ positive semidefinite matrices, $\mathcal{S}_+^n := \{\mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} \succeq \mathbf{0}\}$.

2. Robust counterparts of general uncertain constraint. In this section, we recall useful results from [2] about tractable reformulations of the robust constraint

(2.1) $$f(\bar{\mathbf{a}} + \mathbf{Pz}, \mathbf{x}) \leq 0\ \forall \mathbf{z} \in \mathcal{Z},$$

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

2.1. Case when the constraint is linear in the uncertainty. We first consider a case where f is linear in $\mathbf{a}$, namely $f(\mathbf{a}, \mathbf{x}) = \mathbf{a}^\top \mathbf{x} - b(\mathbf{x})$.

PROPOSITION 2.1. *The decision variable $\mathbf{x}$ satisfies $\bar{\mathbf{a}}^\top \mathbf{x} + \mathbf{z}^\top \mathbf{P}^\top \mathbf{x} \leq b(\mathbf{x})\ \forall \mathbf{z} \in \mathcal{Z}$ if and only if it satisfies*

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

where $\delta^\star(\mathbf{y}|\mathcal{Z}) := \sup_{\mathbf{z}\in\mathcal{Z}} \mathbf{y}^\top \mathbf{z}$ is the so-called support function *of $\mathcal{Z}$ [41, Chapter 13].*

Proposition 2.1 uncovers the role played by the uncertainty set $\mathcal{Z}$ through its support function $\delta^\star(\cdot|\mathcal{Z})$. Though general, Proposition 2.1 is also a practical statement, because the support function can be computed for a wide range of uncertainty sets (see [2, 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 2.1, given the sense of the inequality. We illustrate this point with the following example.

TABLE 3
Examples of uncertainty sets and their associated support function.

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

Example 2.2. If $\mathcal{Z} = \{\mathbf{z} \,:\, \mathbf{D}\mathbf{z} \leq \mathbf{d}\}$ is a polyhedron, then the support function of $\mathcal{Z}$ is given by $\delta^\star(\mathbf{y}|\mathcal{Z}) = \min_{\mathbf{v}\geq\mathbf{0}:\mathbf{D}^\top\mathbf{v}=\mathbf{y}} \mathbf{d}^\top\mathbf{v}$, and according to Proposition 2.1, the robust linear constraint is equivalent to the existence of a feasible vector $\mathbf{v}$, $\mathbf{v} \geq \mathbf{0}$, and $\mathbf{D}^\top\mathbf{v} = \mathbf{y}$ satisfying $\bar{\mathbf{a}}^\top\mathbf{x} + \mathbf{d}^\top\mathbf{v} \leq b(\mathbf{x})$.

2.2. Case when the constraint is concave in the uncertainty. We now consider the case where $f(\mathbf{a}, \mathbf{x})$ is a concave function in $\mathbf{a}$. Indeed, constraint (2.1) is equivalent to

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

and the inner maximization problem is well posed. In this case, [2] provides an equivalent reformulation under the technical assumption that $\bar{\mathbf{a}}$ is *regular*, that is, when $\bar{\mathbf{a}}$ is within the relative interior of the domain of f, $\bar{\mathbf{a}} \in \mathrm{ri}(\mathrm{dom}(f(\cdot, \mathbf{x})))$ in short. From a high-level perspective, they require f to be properly defined in the vicinity of the nominal value $\bar{\mathbf{a}}$. So $\bar{\mathbf{a}}$ 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 2.3. *If* $\bar{\mathbf{a}} \in \mathrm{ri}(\mathrm{dom}(f(\cdot, \mathbf{x})))$ *and* $f(\mathbf{a}, \mathbf{x})$ *is concave in* $\mathbf{a}$, *then* $\mathbf{x}$ *satisfies*

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

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

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

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

Remarkably, Proposition 2.3 provides a robust reformulation of the nonlinear constraints (2.1) where the terms involving $\mathcal{Z}$ are independent from those involving f. In this regard, the auxiliary variable $\mathbf{v}$ plays a critical role: $\mathbf{v}$ is linearly impacted by the uncertainty on $\mathbf{z}$ while being coupled with $\mathbf{x}$ through the concave conjugate $f_\star$. Proposition 2.3 shows that a concave robust constraint in the decision variable $\mathbf{x}$ can be expressed as a *linear* robust constraint in the variables $(\mathbf{x}, \mathbf{v})$. In the special case where $f(\mathbf{a}, \mathbf{x}) = \mathbf{a}^\top\mathbf{x} - b(\mathbf{x})$ is linear in $\mathbf{a}$, we recover Proposition 2.1.

We refer the reader to [2, section 4, Table 4] for examples.

3. Probabilistic guarantees for independent sub-Gaussian uncertainty.
Most probabilistic guarantees in the literature are valid under some light-tail assumption on the random vector $\tilde{\mathbf{z}}$, in particular that $\tilde{\mathbf{z}}$ is bounded almost surely or follows a Gaussian distribution. All our results in this section hold under the weaker assumption that the random vector $\tilde{\mathbf{z}}$ is sub-Gaussian.

DEFINITION 3.1 ([40, Definition 1.2]). *A random variable $\tilde{z} \in \mathbb{R}$ is said to be* sub-Gaussian *with parameter σ^2, denoted $\tilde{z} \sim subG(\sigma^2)$, if $\tilde{z}$ is centered, i.e., $\mathbb{E}[\tilde{z}] = 0$, and $\forall\, s \in \mathbb{R}$, $\mathbb{E}\left[e^{s\tilde{z}}\right] \leq e^{\frac{s^2\sigma^2}{2}}$.*

Many random variables encountered in practice are or can be assumed to be sub-Gaussian. Naturally, centered Gaussian random variables are sub-Gaussian. Unbounded random variables with Gaussian tails are sub-Gaussian (see [44, Theorem 2.6]). Bounded random variables are also a special case of sub-Gaussian random variables, as a consequence of Hoeffding's inequality (see Appendix A, Lemma A.2), irrespective of the distribution on their bounded support (e.g., uniform, triangular). Finally, the image of a random vector $\mathbf{X}$ with independent coordinates by some function f, $f(\mathbf{X})$, is sub-Gaussian if one of the following three conditions holds: (a) f satisfies the bounded difference property [44, Corollary 2.21]; (b) $\mathbf{X}$ is Gaussian and f is a Lipschitz function [44, Theorem 2.26]; (c) $\mathbf{X}$ is bounded and f is a convex and Lipschitz function [44, Theorem 3.24]. The parameter σ^2 is usually referred to as a *variance proxy*, because in the case where $\tilde{z}$ is normally distributed, the inequality above holds with $\sigma^2 = \text{Var}(\tilde{z})$. In the rest of the paper, we make the following assumption on the randomness $\tilde{\mathbf{z}}$.

ASSUMPTION 3.2. *We assume that the coordinates of the uncertain parameter $\tilde{\mathbf{z}} \in \mathbb{R}^L$ are L sub-Gaussian random variables with variance proxy 1.*

Assumption 3.2 holds in particular if the coordinates of $\tilde{\mathbf{z}}$ are (a) Gaussian random variables with mean 0 and variance 1 or (b) centered random variables in $[-1, 1]$, which is a common assumption in the RO literature (see [4, 8], for instance). As summarized in Table 4, common assumptions on the uncertainty parameter $\tilde{\mathbf{z}}$ found in the literature all can be seen as assuming sub-Gaussian random variables for some well-chosen variance proxy. We report some useful properties of sub-Gaussian random variables in Appendix A.

In this section, we also require $\tilde{\mathbf{z}}$ to have independent coordinates.

ASSUMPTION 3.3. *We assume that the coordinates of the uncertain parameter $\tilde{\mathbf{z}} \in \mathbb{R}^L$ are independent.*

In the Gaussian case, for instance, Assumption 3.3 implies that the mean and the covariance of $\tilde{\mathbf{a}}$ are known to the decision maker and equal to $\bar{\mathbf{a}}$ and $\mathbf{P}\mathbf{P}^\top$, respectively. As a result, it is an admittedly strong assumption—although made in many of the previous analysis on probabilistic guarantees for RO. In practice, parameters of the linear model $\tilde{\mathbf{a}} = \bar{\mathbf{a}} + \mathbf{P}\tilde{\mathbf{z}}$ are typically unknown and need to be estimated from data. In section 5, we will depart from this independence assumption and allow for model misspecification—at the expense of weaker probabilistic guarantees.

3.1. A posteriori guarantee when the constraint is linear in a. First, we consider the linear case where $f(\mathbf{a}, \mathbf{x}) = \mathbf{a}^\top \mathbf{x} - b(\mathbf{x})$ and prove a solution-specific probabilistic guarantee.

THEOREM 3.4. *Under Assumptions* 3.2 *and* 3.3*, for any $\mathbf{x} \in \mathcal{X}$ satisfying $\bar{\mathbf{a}}^\top \mathbf{x} +$*

TABLE 4
Summary of relevant assumptions on $\tilde{\mathbf{z}}$ found in the literature and the corresponding variance proxy. Proofs for the last three lines can be found in [3, Chapter 2.4]. If $\tilde{\mathbf{z}}$ has variance proxy σ^2 instead of 1, then the rate $^1/_2$ in the probabilistic guarantee should be replaced by $^1/2\sigma^2$.

Assumption on $\tilde{z}$	Variance proxy
$\mathbb{E}[\tilde{z}] = 0$ $\|\tilde{z}\| \leq 1$, 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 $\|\tilde{z}\| \leq 1$, a.s.	$^1/_3$
$\mathbb{E}[\tilde{z}]$ symmetric w.r.t. 0 $\|\tilde{z}\| \leq 1$, 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 $\|\tilde{z}\| \leq 1$, a.s. $\mathrm{Var}(\tilde{z}) \leq \nu^2 \leq {}^1/_3$	$\min_{c \geq 0} \left\{ c \, : \, \forall t, \ c^2 t^2 \geq 2\ln(3\nu^2 \sinh(t)/t + 1 - 3\nu^2) \right\} \leq 1$

$\delta^\star(\mathbf{P}^\top \mathbf{x} | \mathcal{Z}) \leq b(\mathbf{x})$, *we have*

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

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

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

Let us denote $t := b(\mathbf{x}) - \bar{\mathbf{a}}^\top \mathbf{x} \geq \delta^\star(\mathbf{P}^\top \mathbf{x} | \mathcal{Z})$. If $\mathbf{P}^\top \mathbf{x} \neq \mathbf{0}$, then $\delta^\star(\mathbf{P}^\top \mathbf{x} | \mathcal{Z}) > 0$, since $\mathcal{Z}$ is full dimensional. Under Assumptions 3.2 and 3.3, $\tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x}$ is sub-Gaussian with variance proxy $\|\mathbf{P}^\top \mathbf{x}\|_2^2$ (see Proposition A.4(a)). Hence, we have the following tail bound (Proposition A.3):

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

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

Note that the bound in Theorem 3.4 is a consequence of a tail bound for sub-Gaussian random variables (Proposition A.3). For Gaussian random variables, this bound is sharp up to polynomial-factor corrections (see [44, Exercise 2.2, Mills ratio]). The probabilistic guarantee of constraint violation in Theorem 3.4 depends on the specific solution $\mathbf{x}$. As a result, this bound cannot be used a priori to scale the size of the uncertainty set $\mathcal{Z}$. Yet, given a robust solution $\mathbf{x}$, the bound only involves quantities which can be easily computed and can provide a useful a posteriori guarantee, as we discuss in section 6. From a practical standpoint, the bound involving the slack term $b(\mathbf{x}) - \bar{\mathbf{a}}^\top \mathbf{x}$ is systematically tighter and 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 $\mathbf{P}^\top \mathbf{x}$, 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 bounds.
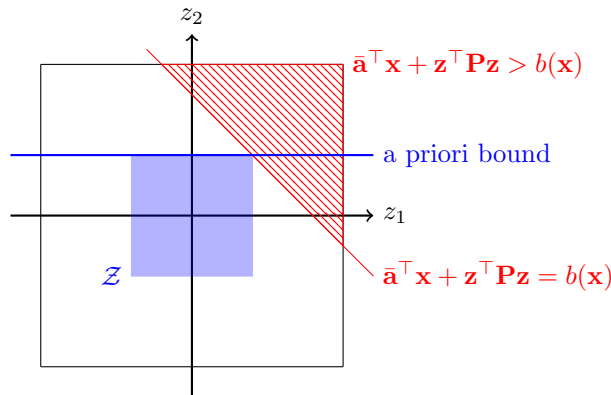
FIG. 1. *Geometrical interpretation of the a posteriori (Theorem 3.4) and a priori (Corollary 3.9) bound in dimension $L = 2$.*

Remark 3.5. Theorem 3.4 can be used to obtain a robust solution $\mathbf{x}$, which would violate the uncertain constraint with probability at most ε by adding the second-order cone constraint,

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

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

Remark 3.6. The first inequality in Theorem 3.4,

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

is valid as long as $\mathbf{x}$ satisfies the linear constraint at $\mathbf{a} = \bar{\mathbf{a}}$: $\bar{\mathbf{a}}^\top \mathbf{x} < b(\mathbf{x})$.

3.2. A priori guarantee and robust complexity. Theorem 3.4 can be extended to give an a priori probabilistic guarantee which does not depend on a specific solution $\mathbf{x}$. Among others, these inequalities are largely used in practice to determine the size of the uncertainty set. Graphically, from Theorem 3.4, every robust solution $\mathbf{x}$ defines a half-space, $\{ \mathbf{z} : \bar{\mathbf{a}}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{P} \mathbf{z} > b(\mathbf{x}) \}$, and the probability of constraint violation corresponds to the area (measured in terms of probability) of this half-space, as depicted in Figure 1. Since $\mathcal{Z} \subseteq \{ \mathbf{z} : \bar{\mathbf{a}}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{P} \mathbf{z} \leq b(\mathbf{x}) \}$, $\{ \mathbf{z} : \bar{\mathbf{a}}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{P} \mathbf{z} = b(\mathbf{x}) \}$ is at worst a supporting hyperplane of $\mathcal{Z}$. Hence, a conservative estimate for the probability of constraint violation can be obtained by taking the worst (in terms of resulting probability) supporting hyperplane. In this section, we rigorously support this geometrical intuition.

We first define the *robust complexity* of a set $\mathcal{Z}$.

DEFINITION 3.7. *For any set $\mathcal{Z} \subseteq \mathbb{R}^L$, we define the* robust complexity *of $\mathcal{Z}$, and denote $\rho(\mathcal{Z})$, the quantity*

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

In statistics, the Rademacher complexity of a set $\mathcal{Z}$ is defined as $\mathcal{R}(\mathcal{Z}) := \mathbb{E}_{\tilde{\mathbf{y}}} \left[\sup_{\mathbf{z} \in \mathcal{Z}} \mathbf{z}^\top \tilde{\mathbf{y}} \right]$, where the coordinates of $\tilde{\mathbf{y}}$ are independently drawn from a

Rademacher distribution, i.e., $\mathbb{P}(\tilde{y}_j = \pm 1) = 1/2$ (see [44, Example 2.25]). This quantity describes the size of the set $\mathcal{Z}$ and drives the so-called uniform law of large numbers (see [44, Chapters 4 and 5]. In this regard, our proposed complexity metric $\rho(\mathcal{Z})$ is a robust analogue of the Rademacher complexity, replacing expectation by worst-case value. Indeed, for any Rademacher vector $\tilde{\mathbf{y}}$, we have $\|\tilde{\mathbf{y}}\|_2 = \sqrt{L}$ 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 $\mathcal{G}(\mathcal{Z}) := \mathbb{E}_{\tilde{\mathbf{y}}}\left[\sup_{\mathbf{z}\in\mathcal{Z}} \mathbf{z}^\top \tilde{\mathbf{y}}\right]$, where the coordinates of $\tilde{\mathbf{y}}$ are independent Gaussian random variables with mean 0 and variance 1 (see [44, Example 2.30 and Chapter 5]). Since $\mathcal{R}(\mathcal{Z}) \leq \sqrt{\frac{\pi}{2}}\mathcal{G}(\mathcal{Z})$ (see [44, Exercise 5.5]), we have $\rho(\mathcal{Z}) \leq \sqrt{\frac{\pi}{2L}}\mathcal{G}(\mathcal{Z})$. Alternatively, by decomposing $\tilde{\mathbf{y}}$ into $\tilde{r}\tilde{\mathbf{u}}$ where $\tilde{\mathbf{u}}$ is a unit vector uniformly 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.

To familiarize ourselves with the notion of robust complexity, we derive some of its basic properties.

PROPOSITION 3.8. *The robust complexity of a set satisfies the following properties:*
(a) *Monotonicity: For any sets $\mathcal{Z}, \mathcal{Z}'$, if $\mathcal{Z}' \subseteq \mathcal{Z}$, then $\rho(\mathcal{Z}') \leq \rho(\mathcal{Z})$.*
(b) *Positive homogeneity: For any set $\mathcal{Z}$ and $\alpha > 0$, $\rho(\alpha\mathcal{Z}) = \alpha\rho(\mathcal{Z})$.*
(c) *Invariance by orthogonal transformation: For any orthogonal matrix $\mathbf{U}$ and any set $\mathcal{Z}$, $\rho(\mathbf{U}\mathcal{Z}) = \rho(\mathcal{Z})$, with $\mathbf{U}\mathcal{Z} := \{\mathbf{U}\mathbf{z} : \mathbf{z} \in \mathcal{Z}\}$.*
(d) *Lipschitz continuity: The robust complexity is 1-Lipschitz continuous with respect to the Hausdorff distance [19], i.e., for any convex sets $\mathcal{Z}, \mathcal{Z}'$, $|\rho(\mathcal{Z}) - \rho(\mathcal{Z}')| \leq d_H(\mathcal{Z}, \mathcal{Z}')$.*

Proof. (a) For any $\mathbf{y}$, $\mathcal{Z}' \subseteq \mathcal{Z}$ yields $\delta^\star(\mathbf{y}|\mathcal{Z}') \leq \delta^\star(\mathbf{y}|\mathcal{Z})$.
(b) For any $\alpha > 0$,

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

(c) The matrix $\mathbf{U}$ being orthogonal, the change of variable $\mathbf{y} = \mathbf{U}\mathbf{y}'$ satisfies $\|\mathbf{y}\|_2^2 = \|\mathbf{y}'\|_2^2$, and

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

(d) The Hausdorff distance between two sets $\mathcal{Z}, \mathcal{Z}'$ can be expressed in terms of their respective support functions as $d_H(\mathcal{Z}, \mathcal{Z}') = \sup_{\mathbf{y}:\|\mathbf{y}\|_2=1} |\delta^\star(\mathbf{y}|\mathcal{Z}) - \delta^\star(\mathbf{y}|\mathcal{Z}')|$. The result follows. $\qquad\square$

Observe that monotonicity and homogeneity are also properties satisfied by the Rademacher and Gaussian complexities. Invariance by orthogonal transformation, however, is satisfied by the Gaussian complexity but not by the Rademacher complexity since the transformed random variable $\tilde{\mathbf{y}}' = \mathbf{U}^{-1}\tilde{\mathbf{y}}$ is not a Rademacher random variable.

Relevant to our analysis, the robust complexity is *the* key quantity controlling a priori probabilistic guarantees.

COROLLARY 3.9. *Under Assumptions* 3.2 *and* 3.3, *for any* $\mathbf{x} \in \mathcal{X}$ *satisfying* $\bar{\mathbf{a}}^\top \mathbf{x} + \delta^\star(\mathbf{P}^\top \mathbf{x} | \mathcal{Z}) \leq b(\mathbf{x})$, *we have*

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

Proof. Taking the worst over all feasible $\mathbf{x}$ in the right-hand side of Theorem 3.4 yields

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

Then,

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

In section 4, we analyze properties of the robust complexity and provide explicit analytic expressions of $\rho(\mathcal{Z})$ for uncertainty sets found in the literature, such as the budget [8] and the box-ellipsoidal [4] uncertainty sets, which hence unify most of the known probabilistic guarantees and their proofs.

3.3. Case when the constraint is concave in the uncertainty. According to Proposition 2.3, a robust constraint that depends on the uncertainty in a concave manner, as in (2.1), can be seen as a linear constraint in the new decision variables $(\mathbf{x}, \mathbf{v})$. Correspondingly, we can now state and derive an analogue of Theorem 3.4 for this general case.

THEOREM 3.10. *Under Assumptions* 3.2 *and* 3.3, *for any* $\mathbf{x} \in \mathcal{X}$ *and* $\mathbf{v}$ *satisfying* $\bar{\mathbf{a}}^\top \mathbf{v} + \delta^\star(\mathbf{P}^\top \mathbf{v} | \mathcal{Z}) - f_\star(\mathbf{v}, \mathbf{x}) \leq 0$, *we have*

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

A formal proof of Theorem 3.10 is given in Appendix B.1. We graphically explain the intuition behind the result in dimension $L = 2$ in Figure 2. Assume $(\tilde{z}_1, \tilde{z}_2) \in [-1, 1]^2$ (black squared box), which is a special case of Assumption 3.2. The shaded blue region corresponds to the uncertainty set $\mathcal{Z}$. Let $\mathbf{x}$ be a robust solution. It induces a constraint on $\mathbf{z}$, $f(\bar{\mathbf{a}} + \mathbf{P}\mathbf{z}, \mathbf{x}) \leq 0$, which is satisfied by all $\mathbf{z} \in \mathcal{Z}$. By concavity, the region $\{\mathbf{z} : f(\bar{\mathbf{a}} + \mathbf{P}\mathbf{z}, \mathbf{x}) > 0\}$ admits a supporting hyperplane (in red) and is contained within a halfspace. Theorem 3.10 provides an explicit description of this halfspace, $\{\mathbf{z} : \bar{\mathbf{a}}^\top \mathbf{x} + \mathbf{v}^\top \mathbf{P}\mathbf{z} - f_\star(\mathbf{v}, \mathbf{x}) > 0\}$.

Theorem 3.10 displays exactly the same probabilistic guarantee as in the linear case except that the bound now involves the extra variable $\mathbf{v}$ instead of $\mathbf{x}$. Recall that in the special case where f is linear, the conjugate $f_\star$ enforces $\mathbf{v} = \mathbf{x}$, and we recover Theorem 3.4. In the general case, the variable $\mathbf{v}$ is introduced to express the robust constraint in a tractable manner, so solving the robust optimization problem provides a vector $\mathbf{v}$ alongside a vector $\mathbf{x}$, and our bound can be computed a posteriori. Taking the worst case over all potential vectors $\mathbf{v}$, we get exactly the same a priori bounds as in the linear case.

COROLLARY 3.11. *Under Assumptions* 3.2 *and* 3.3, *for any* $\mathbf{x} \in \mathcal{X}$ *and* $\mathbf{v}$ *satisfying* $\bar{\mathbf{a}}^\top \mathbf{v} + \delta^\star(\mathbf{P}^\top \mathbf{v} | \mathcal{Z}) - f_\star(\mathbf{v}, \mathbf{x}) \leq 0$, *we have*

$$\mathbb{P}\left(f(\bar{\mathbf{a}} + \mathbf{P}\tilde{\mathbf{z}}, \mathbf{x}) > 0\right) \leq \exp\left(-\tfrac{1}{2}\rho(\mathcal{Z})^2\right).$$
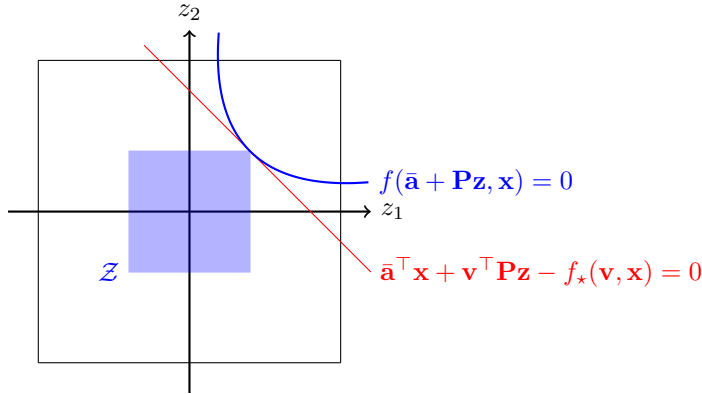
FIG. 2. *Geometrical proof of Theorem* 3.10 *in dimension $L = 2$. (See online version for color.)*

4. Robust complexity of a set. The previous section elicited the central role of the robust complexity of the uncertainty set, $\rho(\mathcal{Z})$, in obtaining a priori probabilistic guarantees. In section 4.1, we derive a closed-form expression for the robust complexity of norm balls and analyze the robust complexity of their intersection and Minkowski sum in section 4.2 and 4.3, respectively. By doing so, we recover most existing results found in the literature. Finally, we provide a geometric interpretation of the robust complexity in section 4.4, which in turns leads to an analytical expression for polyhedral uncertainty sets and an optimization problem to evaluate $\rho(\mathcal{Z})$ numerically in general. 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 4.1. $\mathcal{Z} = \left\{ \mathbf{z} \in \mathbb{R}^L : \|\mathbf{z}\|_p \leq \Gamma \right\}$ *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

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

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

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. Equivalently, in order for the constraint to be violated with probability at most ε, it suffices to take $\Gamma \geq \frac{1}{\kappa(p)} \sqrt{2 \ln(1/\varepsilon)}$.

Proposition 4.1 applies to norm balls which are isotropic, i.e., 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 (see the proof in Appendix B.2).

PROPOSITION 4.2. *Let* $\mathbf{\Lambda} = Diag(\lambda)$ *be a diagonal matrix with* $\lambda > \mathbf{0}$*, and let* $p \in [1, \infty]$. *Then the set* $\mathcal{Z} = \left\{ \mathbf{z} \in \mathbb{R}^L : \|\mathbf{\Lambda z}\|_p = (\sum_i |\lambda_i z_i|^p)^{1/p} \leq 1 \right\}$ *has a robust complexity of*

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

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

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

PROPOSITION 4.3. *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 Assumptions 3.2 and 3.3,* $\tilde{\mathbf{z}} \in \mathcal{Z}_1$ *almost surely, then the probabilistic guarantee is driven by* $\rho(\mathcal{Z}_2)$.

Proposition 4.3(a) is nontrivial. 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 $\mathcal{Z}_2$. Yet, Proposition 4.3(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 4.3(b) improves upon Proposition 4.3(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 $ri(\mathcal{Z}_1) \cap ri(\mathcal{Z}_2) \neq \emptyset$, we have [2, Lemma 6.4] $\delta^\star(\mathbf{y}|\mathcal{Z}) = \min_{\mathbf{v}} \delta^\star(\mathbf{v}|\mathcal{Z}_1) + \delta^\star(\mathbf{y} - \mathbf{v}|\mathcal{Z}_2)$. In addition, by definition of the robust complexity, $\delta^\star(\cdot|\mathcal{Z}_i) \geq \rho(\mathcal{Z}_i)\|\cdot\|_2$ so that

$$\delta^\star(\mathbf{y}|\mathcal{Z}) \geq \min_{i=1,2} \rho(\mathcal{Z}_i) \min_{\mathbf{v}} [\|\mathbf{v}\|_2 + \|\mathbf{y} - \mathbf{v}\|_2] \quad \geq \min_{i=1,2} \rho(\mathcal{Z}_i) \|\mathbf{y}\|_2,$$

where the last inequality follows from the triangle inequality.

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

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

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

Remark 4.4. The proof of Proposition 4.3(b) is informative for an a posteriori bound as well. Indeed, we have

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

which might be tighter than the bound from Theorem 3.4.

Example 4.5. *The budget uncertainty set.* The budget uncertainty set defined as

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

with $1 \leq \rho \leq L$, and introduced by [8], is the intersection of the ℓ_∞ unit-ball with an ℓ_1-ball. The robust constraint $\bar{\mathbf{a}}^\top\mathbf{x} + \mathbf{z}^\top\mathbf{P}^\top\mathbf{x} \leq b \;\forall\mathbf{z} \in \mathcal{Z}_{\ell_\infty \cap \ell_1}$ is then equivalent to

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

requiring the introduction of the L new variables $\mathbf{v}$, 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 $\mathbf{a}$, there is a total of $(2L+1)m$ new variables and $4Lm$ new constraints needed. According to Proposition 4.3(b), $\mathcal{Z}_{\ell_\infty \cap \ell_1}$ induces a probabilistic guarantee of $\exp\left(-\rho^2/2L\right)$, recovering the original result from [8]. Our result holds whenever $\tilde{\mathbf{z}}$ is sub-Gaussian with variance proxy 1 (Assumption 3.2) and $\|\tilde{\mathbf{z}}\|_\infty \leq 1$ but does not require $\tilde{\mathbf{z}}$ to be symmetrically distributed. Similarly, for the so-called box-ellipsoidal uncertainty set from [4], $\mathcal{Z}_{\ell_\infty \cap \ell_2} := \left\{\mathbf{z} \in \mathbb{R}^L : \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_2 \leq \rho\right\}$, we recover their $e^{-\rho^2/2}$ guarantee.

4.3. Minkowski sum of 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 4.6. *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 $\delta^\star(\mathbf{y}|\mathcal{Z}) = \delta^\star(\mathbf{y}|\mathcal{Z}_1) + \delta^\star(\mathbf{y}|\mathcal{Z}_2)$ [2, Lemma 6.3]. Hence,

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

Taking the minimum over all $\mathbf{y}$ such that $\|\mathbf{y}\|_2 = 1$ concludes the proof. $\square$
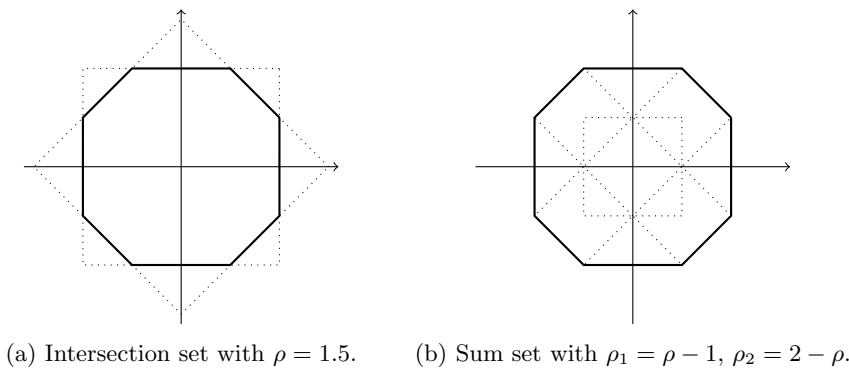
Example 4.7. *Alternative to the budget uncertainty set.* From a modeling perspective, the budget uncertainty $\mathcal{Z}_{\ell_\infty \cap \ell_1}$ set is geometrically very similar to the $\ell_\infty + \ell_1$ set, defined as

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

In dimension $L = 2$, with proper scaling, those two sets are indeed identical as shown in Figure 3, while this is no longer the case in a higher dimension (see Figure 4 for $L = 3$).

In terms of tractability, however, the sum is preferable over the intersection because it does not require additional variables $\mathbf{v}$. The robust constraint $\bar{\mathbf{a}}^\top\mathbf{x} + \mathbf{z}^\top\mathbf{P}^\top\mathbf{x} \leq b \;\forall\mathbf{z} \in \mathcal{Z}_{\ell_\infty + \ell_1}$ is equivalent to

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

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

FIG. 3. *Comparison of the intersection and sum set in dimension $L = 2$.*

(a) Intersection set with $\rho = 2.5$.   (b) Sum set with $\rho_1 = (\rho - 1)/2$, $\rho_2 = (3 - \rho)/2$.
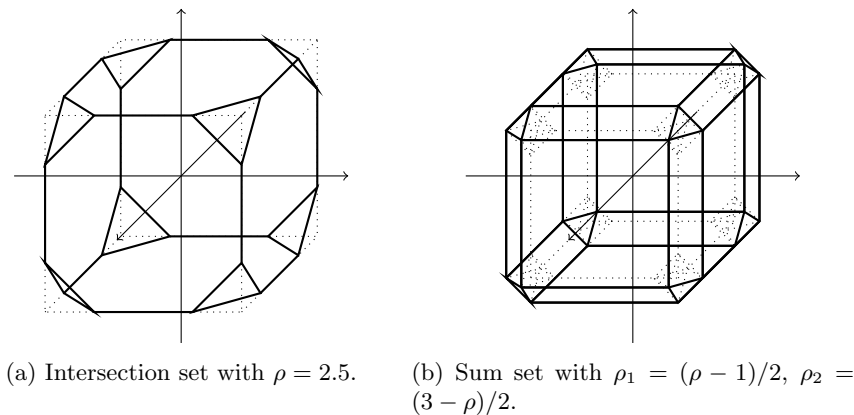
FIG. 4. *Comparison of the intersection and sum set in dimension $L = 3$.*

which requires $L + 1$ extra variables and $3L$ extra constraints to linearize the ℓ_1- and ℓ_∞-norms of $\mathbf{P}^\top \mathbf{x}$. 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 $\mathbf{a}$, we need only a total of $(L + 1)$ new variables and $3L$ new constraints. Applying Proposition 4.6, we can show that the $\mathcal{Z}_{\ell_\infty + \ell_1}$ set yields a probabilistic guarantee of $\exp(-\frac{1}{2} \{\rho_1 + \rho_2 L^{-1/2}\}^2)$. 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 in a numerical example in section 6.3.

Example 4.8. *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

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

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

4.6, it yields a probabilistic guarantee of $\exp(-\frac{1}{2}\{\rho_1 + \rho_2\}^2)$, which is independent of the number of uncertain parameters L.

Remark 4.9. Sets of the form $\mathcal{Z}_{\ell_\infty + \ell_2}$ have been consider by [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 $\tilde{\mathbf{z}}$ 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 $\tilde{\mathbf{z}}$ can be modeled by decomposing $\tilde{\mathbf{z}}$ into $\tilde{\mathbf{z}}_1 + \tilde{\mathbf{z}}_2$, where $\tilde{\mathbf{z}}_1$ and $\tilde{\mathbf{z}}_2$ are unknown parameters with mean $\mathbf{0}$. Intuitively, given $\tilde{\mathbf{z}}_1$, $\tilde{\mathbf{z}}$ has mean $\tilde{\mathbf{z}}_1 \neq \mathbf{0}$, and $\tilde{\mathbf{z}}_2$ 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. Monotonicity of the robust complexity (Proposition 3.8(a)) 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 $\mathcal{B}(\mathbf{0}, \mathbf{r}) = \{\mathbf{z} : \|\mathbf{z}\|_2 \leq r\}$.

PROPOSITION 4.10. *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., $\rho(\mathcal{Z}) = \max_{r \geq 0} \{r : \mathcal{B}(\mathbf{0}, \mathbf{r}) \subseteq \mathcal{Z}\}$.*

Proof. Denote by $r^\star$ 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, $\mathcal{B}(\mathbf{0}, \mathbf{r}^\star) \subseteq \mathcal{Z}$, so $\rho(\mathcal{Z}) \geq \rho(\mathcal{B}(\mathbf{0}, \mathbf{r}^\star)) = r^\star$, where the last equality follows from Proposition 4.1. We consider a vector $\mathbf{z}^\star$, $\|\mathbf{z}^\star\|_2 = r^\star$, which lies on the boundary of $\mathcal{Z}$. Such a vector exists by optimality of $r^\star$. Then, $\mathbf{z}^\star$ defines a hyperplane that is tangent to $\mathcal{Z}$, and hence proves that the linear optimization problem $\max_{\mathbf{z} \in \mathcal{Z}} \mathbf{z}^\top \mathbf{z}^\star$ admits $\mathbf{z} = \mathbf{z}^\star$ as an optimal solution. As a result, by considering $\mathbf{y} = \mathbf{z}^\star / r^\star$ we have $\rho(\mathcal{Z}) \leq r^\star$. □

The optimization problem in Proposition 4.10 is a special case of the maximum inscribed ellipsoid [50] or the maximum inscribed sphere problem [46]. Although we fix the center of the ellipsoid to $\mathbf{0}$ and only allow for shape matrices of the form $r\mathbf{I}_L$, solving this optimization problem can be challenging in general. For polyhedral uncertainty sets defined with k linear constraints, $\mathcal{Z} = \{\mathbf{z} : \mathbf{Dz} \leq \mathbf{d}\}$, however, it can be solved in closed-form as follows:

$$\rho(\mathcal{Z}) = \min_{i=1,\ldots,k} \frac{d_i}{\|\mathbf{D}^\top \mathbf{e}_i\|_2}.$$

5. Probabilistic guarantees for dependent uncertainty. In the preceding section, we leveraged the assumption that coordinates of $\tilde{\mathbf{z}}$ are independent (Assumption 3.3) to derive our probabilistic guarantees. We now discuss how this assumption can be relaxed and its impact on the resulting guarantees. In section 5.1, we show that exponential, yet weaker, bounds can be derived when $\tilde{\mathbf{z}}$ is sub-Gaussian (Assumption 3.2), but its coordinates are not necessarily independent. In section 5.2, we depart from both Assumptions 3.2 and 3.3 and derive polynomial probabilistic guarantees under first and second moment conditions.

5.1. Exponential guarantees for sub-Gaussian random variables. From a high-level perspective, the proofs of Theorems 3.4 and 3.10 are essentially a tail-

TABLE 5

*Valid lower bound on $\rho_1(\mathcal{Z}) := \min_{\mathbf{y}:\|\mathbf{y}\|_1=1} \max_{\mathbf{z}\in\mathcal{Z}} \mathbf{y}^\top \mathbf{z}$, for some common uncertainty sets. For each set, we also provide sufficient conditions (s.c.) to obtain a probabilistic guarantee of at least ε, i.e., $\rho_1(\mathcal{Z}) \geq \gamma(\varepsilon) := \sqrt{2\ln(1/\varepsilon)}$. Instances denoted by a * are valid under the assumption that the true uncertain parameter $\tilde{\mathbf{z}}$ satisfies $\|\tilde{\mathbf{z}}\|_\infty \leq 1$.*

Uncertainty set	Definition	$\rho_1(\mathcal{Z}) \geq \ldots$	s.c. for ε-guarantee
Norm-set	$\{\mathbf{z} : \|\mathbf{z}\|_p \leq \Gamma\}$	$\Gamma L^{-1/p}$	$\Gamma \geq L^{1/p}\gamma(\varepsilon)$
Budget set*	$\{\mathbf{z} : \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_1 \leq \Gamma\}$	Γ/L	$\Gamma \geq L\gamma(\varepsilon)$
Box-ellipsoidal set*	$\{\mathbf{z} : \|\mathbf{z}\|_\infty \leq 1, \|\mathbf{z}\|_2 \leq \Gamma\}$	$\Gamma/\sqrt{L}$	$\Gamma \geq \sqrt{L}\gamma(\varepsilon)$
$\ell_\infty+\ell_1$ set	$\{\mathbf{z}_1 + \mathbf{z}_2 : \|\mathbf{z}_1\|_\infty \leq \rho_1, \|\mathbf{z}_2\|_1 \leq \rho_2\}$	$\rho_1 + \rho_2/L$	$\rho_1 \geq \alpha\gamma(\varepsilon),\ \alpha \in [0,1]$ $\rho_2 \geq (1-\alpha)L\gamma(\varepsilon)$
$\ell_\infty+\ell_2$ set	$\{\mathbf{z}_1 + \mathbf{z}_2 : \|\mathbf{z}_1\|_\infty \leq \rho_1, \|\mathbf{z}_2\|_2 \leq \rho_2\}$	$\rho_1 + \rho_2/\sqrt{L}$	$\rho_1 \geq \alpha\gamma(\varepsilon),\ \alpha \in [0,1]$ $\rho_2 \geq (1-\alpha)\sqrt{L}\gamma(\varepsilon)$

bound $\mathbb{P}\left(\tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x} > t\right)$ for some $t \geq \delta^\star(\mathbf{P}^\top \mathbf{x}|\mathcal{Z}) > 0$. When the coordinates of $\tilde{\mathbf{z}}$ are L independent sub-Gaussian random variables with parameter 1, the random variable $\tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x}$ is itself sub-Gaussian, with parameter $\|\mathbf{P}^\top \mathbf{x}\|_2^2$. Without independence, $\tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x}$ remains sub-Gaussian (Proposition A.4(b)), with a higher parameter, however. We formalize this line of proof in the following theorem.

THEOREM 5.1. *Under Assumption 3.2, for any $\mathbf{x} \in \mathcal{X}$ satisfying $\bar{\mathbf{a}}^\top \mathbf{x}+\delta^\star(\mathbf{P}^\top \mathbf{x}|\mathcal{Z}) \leq b(\mathbf{x})$, we have*

$$\mathbb{P}\left(\bar{\mathbf{a}}^\top \mathbf{x} + \tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x} > b(\mathbf{x})\right) \leq \exp\left(-\frac{(b(\mathbf{x}) - \bar{\mathbf{a}}^\top \mathbf{x})^2}{2\|\mathbf{P}^\top \mathbf{x}\|_1^2}\right) \leq \exp\left(-\frac{\delta^\star(\mathbf{P}^\top \mathbf{x}|\mathcal{Z})^2}{2\|\mathbf{P}^\top \mathbf{x}\|_1^2}\right).$$

Proof. We follow the proof of Theorem 3.4. Under Assumption 3.2, $\tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x}$ is sub-Gaussian with variance proxy $\|\mathbf{P}^\top \mathbf{x}\|_1^2$ (see Proposition A.4(b)). Hence, we have the following tail bound (Proposition A.3):

$$\mathbb{P}\left(\tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x} > t\right) \leq \exp\left(-\frac{t^2}{2\|\mathbf{P}^\top \mathbf{x}\|_1^2}\right). \qquad \square$$

Since $\|\mathbf{P}^\top \mathbf{x}\|_2 \leq \|\mathbf{P}^\top \mathbf{x}\|_1$, the bound obtained from relaxing Assumption 3.3 is weaker than the one from Theorem 3.4 but shares a similar flavor. From the a posteriori bound, we derive an a priori probabilistic guarantee (the proof is omitted).

COROLLARY 5.2. *Under Assumptions 3.2, for any $\mathbf{x} \in \mathcal{X}$ satisfying $\bar{\mathbf{a}}^\top \mathbf{x}+ \delta^\star(\mathbf{P}^\top \mathbf{x}|\mathcal{Z}) \leq b(\mathbf{x})$, we have*

$$\mathbb{P}\left(\bar{\mathbf{a}}^\top \mathbf{x} + \tilde{\mathbf{z}}^\top \mathbf{P}^\top \mathbf{x} > b(\mathbf{x})\right) \leq \exp\left(-\tfrac{1}{2}\rho_1\left(\mathcal{Z}\right)^2\right),$$

where $\rho_1\left(\mathcal{Z}\right) = \min_{\mathbf{y}:\|\mathbf{y}\|_1=1} \delta^\star(\mathbf{y}|\mathcal{Z})$.

This a priori probabilistic guarantee involves the quantity $\rho_1\left(\mathcal{Z}\right) = \min_{\mathbf{y}:\|\mathbf{y}\|_1=1} \delta^\star(\mathbf{y}|\mathcal{Z})$ that is almost identical to the robust complexity $\rho(\mathcal{Z})$ except that $\mathbf{y}$ has unit ℓ_1-norm instead of ℓ_2-norm. In particular, we have $\rho(\mathcal{Z})/\sqrt{L} \leq \rho_1(\mathcal{Z}) \leq \rho(\mathcal{Z})$, and can provide closed-form expressions for the most widely used uncertainty sets by conducting an analysis similar to that in section 4 (see Table 5).

5.2. Polynomial guarantees for bounded covariance matrix. In this section, we depart from the sub-Gaussian assumption and provide polynomial probabilistic guarantees under some restrictions on the first two moments of $\tilde{\mathbf{z}}$, as in [7].

ASSUMPTION 5.3. *We assume that the coordinates of the uncertain parameter $\tilde{\mathbf{z}} \in \mathbb{R}^L$ are L random variables with zero mean and covariance matrix bounded by some semidefinite positive matrix $\mathbf{\Sigma} \succeq \mathbf{0}$, i.e., $\mathbb{E}[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^\top] \preceq \mathbf{\Sigma}$.*

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

THEOREM 5.4. *Under Assumption 5.3, for any $\mathbf{x} \in \mathcal{X}$ satisfying $\bar{\mathbf{a}}^\top\mathbf{x} + \delta^\star(\mathbf{P}^\top\mathbf{x}|\mathcal{Z}) \leq b(\mathbf{x})$, we have*

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

Proof. We follow the proof of Theorem 3.4. Instead of a Chernoff bound, we apply a tight version of Chebyshev's inequality due to [37],

$$\mathbb{P}\left(\tilde{\mathbf{z}}^\top\mathbf{P}^\top\mathbf{x} > t\right) \leq \left[1 + \frac{t^2}{\|\mathbf{\Sigma}^{1/2}\mathbf{P}^\top\mathbf{x}\|_2^2}\right]^{-1},$$

with $t = b(\mathbf{x}) - \bar{\mathbf{a}}^\top\mathbf{x} \geq \delta^\star(\mathbf{P}^\top\mathbf{x}|\mathcal{Z}) \geq 0$. $\square$

Remark 5.5. Since $[1 + 1/t]^{-1} \leq t$ for $t \geq 0$, the bound in Theorem 5.4 is always tighter than Chebyshev's. Actually, [37] exhibits distributions for which the bound is tight. Consequently, this bound could be (and has been) used to derive equivalent reformulations of ambiguous chance constraints with first- and second-order information [47, 30].

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

PROPOSITION 5.6. *Under Assumption 5.3, for any $\mathbf{x} \in \mathcal{X}$ satisfying $\bar{\mathbf{a}}^\top\mathbf{x} + \delta^\star(\mathbf{P}^\top\mathbf{x}|\mathcal{Z}) \leq b(\mathbf{x})$, we have*

$$\mathbb{P}\left(\bar{\mathbf{a}}^\top\mathbf{x} + \tilde{\mathbf{z}}^\top\mathbf{P}^\top\mathbf{x} > b(\mathbf{x})\right) \leq \left[1 + \rho(\mathcal{Z})^2/\lambda_{\max}(\mathbf{\Sigma})\right]^{-1}.$$

We omit the proof for brevity. 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 $f(\mathbf{a}, \mathbf{x})$ is a concave function of $\mathbf{a}$.

6. Numerical experiments on a facility location example. In this section, we illustrate how our results materialize on a facility location example. In particular, some decision variables are binary, and multiple constraints are subject to uncertainty. In this context, a posteriori bounds are significantly tighter due to discreteness of $\mathbf{x}$ and hence are especially valuable when considering the probability of multiple constraints being violated. We also implement and compare uncertainty regions described as intersections and sums of norm balls.

6.1. Problem formulation. Given a set of n potential facilities and m customers, the facility location problem consists of constructing facilities $i = 1, \ldots, n$ at

cost c_i in order to satisfy demand at minimal cost, i.e., solve

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

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 is 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 "$\mathbf{a}^\top \mathbf{x} \le b$,"

(6.1)
$$\begin{pmatrix} -u_i \\ \mathbf{d} \end{pmatrix}^\top \begin{pmatrix} x_i \\ \mathbf{X}\mathbf{e}_i \end{pmatrix} \le 0 \; \forall i = 1,\dots,n,$$

which we want to protect against uncertainty in the demand vector $\mathbf{d}$. Values for the nominal problem are taken from the p1 instance of [31] with $n = 10$ facilities and $m = 50$ customers. As in [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 $\tilde{\mathbf{z}}$ satisfies Assumptions 3.2 and 3.3. For simulation purposes, we will consider three particular distributions for $\tilde{\mathbf{z}}$:
- *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

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

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

6.2. Box uncertainty set. We first consider a box uncertainty set

$$\mathcal{Z}_\Gamma^{\mathrm{box}} = \{\mathbf{z} \in \mathbb{R}^m \; : \; \|\mathbf{z}\|_\infty \le \Gamma\},$$

as in [1]. According to Corollary 3.9, the uncertainty set $\mathcal{Z}_\Gamma^{\mathrm{box}}$ a priori induces a probabilistic guarantee of $\exp(-\rho(\mathcal{Z}_\Gamma^{\mathrm{box}})^2/2)$ with $\rho(\mathcal{Z}_\Gamma^{\mathrm{box}}) = \Gamma$ (Proposition 4.1). As for posteriori bounds, given a robust solution $(\mathbf{x}, \mathbf{X})$, we have a probabilistic guarantee of

$$\exp\left(-\frac{1}{2} \frac{|\mathrm{Diag}(\bar{\mathbf{d}})^\top \mathbf{X}\mathbf{e}_i - u_i x_i|^2}{\epsilon_0^2 \|\mathrm{Diag}(\bar{\mathbf{d}})^\top \mathbf{X}\mathbf{e}_i\|_2^2} \right).$$
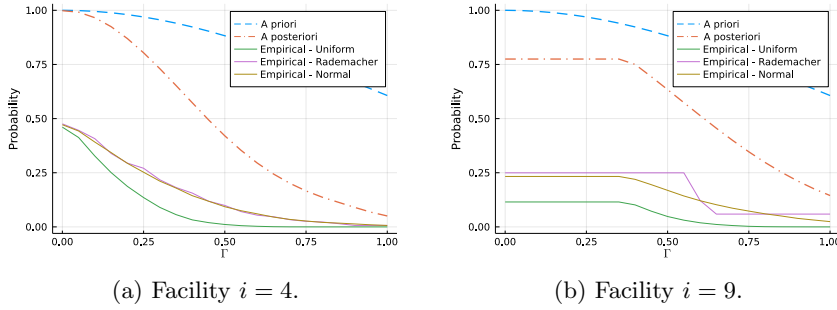
(a) Facility $i = 4$.   (b) Facility $i = 9$.

FIG. 5. *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.*
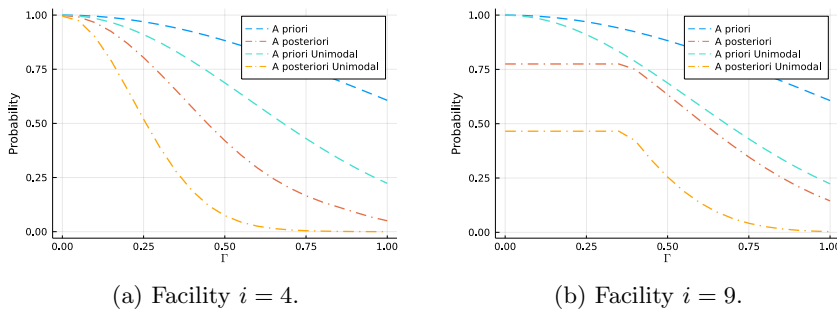

(a) Facility $i = 4$.   (b) Facility $i = 9$.

FIG. 6. *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.*

Figure 5 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 $\tilde{\mathbf{z}}$. For instance, if we assume that $\tilde{z}_j$ admits a bounded symmetric unimodal distribution (such as the uniform distribution), then, as reported in Table 4, $\mathcal{Z}_\Gamma^{\mathrm{box}}$ 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 of 3 as well. Figure 6 compares the a priori and a posteriori guarantees with and without this assumption for the two previous constraints.

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

$$\mathcal{Z}_\Gamma^{\mathrm{budget}} = \left\{ \mathbf{z} \in \mathbb{R}^m \ : \ \|\mathbf{z}\|_\infty \leq 1, \ \|\mathbf{z}\|_1 \leq \Gamma \right\},$$
$$\mathcal{Z}_\Gamma^{\mathrm{sum}} = \left\{ \mathbf{z}_1 + \mathbf{z}_2 \in \mathbb{R}^m \ : \ \|\mathbf{z}_1\|_\infty \leq \Gamma_1, \ \|\mathbf{z}_2\|_1 \leq \Gamma_2 \right\},$$

in terms of a priori and a posteriori guarantees, as discussed in section 4.2. To derive useful conclusions, probabilistic guarantees need to be viewed in contrast to the conservatism of the corresponding solutions. Figure 7 represents the trade-off
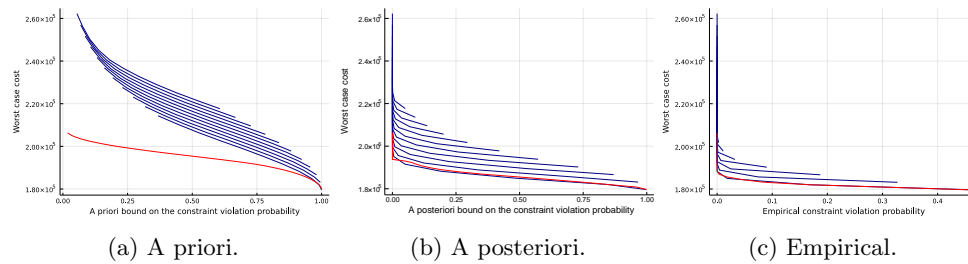
(a) A priori.   (b) A posteriori.   (c) Empirical.

FIG. 7. *Comparison of the trade-off worst-case cost vs. a priori (left) and a posteriori probabilistic guarantee (middle) and an 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. *(See online version for color.)*

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 nearly 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. Because of the absence of theoretical guarantees for uncertainty sets defined as Minkowski sums, these comparisons were previously unavailable. Our experiments illustrate that sum sets can be a viable alternative to budget uncertainty sets and provide a comparable conservatism-risk trade-off at the expense of being parametrized by two hyperparameters. In addition, as previously discussed, their robust counterpart is more tractable, especially when multiple constraints are affected by uncertainty. In our experiments, solve times for the sum uncertainty set were two orders of magnitude smaller than for the budget uncertainty set: Table 6 reports summary statistics on computational time required to solve one instance of the robust facility location problem with budget vs. sum sets, using the commercial solver `Gurobi v.9.1.1` on an Intel i7@3.1GHz processor with 16GB of RAM. Since the sum uncertainty set involves two hyperparameters, while $\mathcal{Z}_{\Gamma}^{\text{budget}}$ involves only one, we had to solve more instances of the robust optimization problem with a sum uncertainty set to generate Figure 7.

TABLE 6

Summary statistics of the solve time (in seconds) to solve one instance of the robust facility location problem for the budget and sum uncertainty set. The different instances correspond to the different values of the budget of uncertainty (Γ for the budget, (Γ_1, Γ_2) for the sum uncertainty set) used to generate Figure 7.

Uncertainty set	# instances	Min	Median	Average	Max
Budget	41	0.51	4.82	4.99	6.57
Sum	121	0.0041	0.0137	0.0160	0.0655

6.4. Extension to multiple constraints. So far, we only looked at the probability of one constraint being violated. To extend our analysis to multiple constraints, we could use a simple union bound technique, alternatively called the Bonferroni
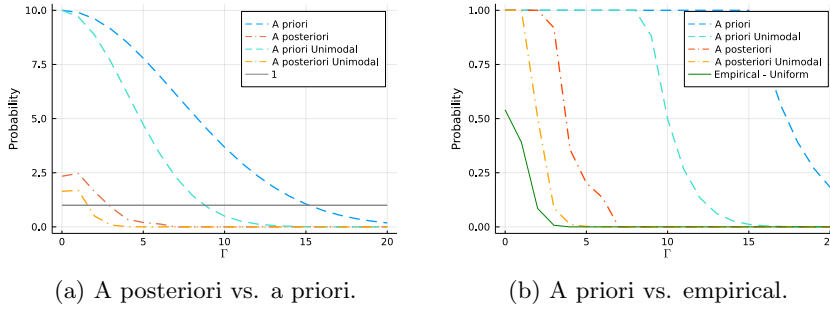
(a) A posteriori vs. a priori.   (b) A priori vs. empirical.

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.*

inequality [10], to 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

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

We apply this reasoning to the $k = n = 10$ production capacity constraints (6.1) for the budget uncertainty set $\mathcal{Z}_\Gamma^{\text{budget}}$. We compute a priori and a posteriori guarantees in Figure 8(a), with and without the assumption that $\tilde{\mathbf{z}}$ is unimodal with respect to $\mathbf{0}$. For small values of Γ, these bounds, and a priori bounds especially, are uninformative because 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 strictly positive. We further compare our bounds (capped at 1) to empirical probabilities (Figure 8(b)). 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 (6.1) similarly computes the worst case for each constraint independently.

6.5. Numerical bounds for dependent uncertainty. Finally, we numerically assess the quality of the bounds when the coordinates of $\tilde{\mathbf{z}}$ are no longer assumed to be independent, as in section 5.1. In particular, we sample $\tilde{\mathbf{z}}$ from a multivariate normal distribution with mean $\mathbf{0}$ and a spiked identity covariance matrix $\mathbf{\Sigma}(\theta) = (1 - \theta)\mathbf{I}_L + \theta \mathbf{e}\mathbf{e}^\top$. Among others, the eigenvalues of $\mathbf{\Sigma}(\theta)$ are within $1 - \theta$ and $1 + \theta(L - 1)$. We consider an ellipsoidal uncertainty set $\mathcal{Z}_\Gamma^{\text{ell}} = \{\mathbf{z} \in \mathbb{R}^m \ : \ \|\mathbf{z}\|_2 \leq \Gamma\}$, which, according to Corollary 5.2 and Table 5, yields a probabilistic guarantee of $\exp(-\Gamma^2/2L)$. Figure 9 compares these a priori bounds to a posteriori ones from Theorem 5.1 and empirical probability of constraint violation. For this setting, we believe a Bernstein-type inequality could provide exponential probabilistic guarantees that depend explicitly on the level of correlation between coordinates and would constitute an interesting extension of our results.
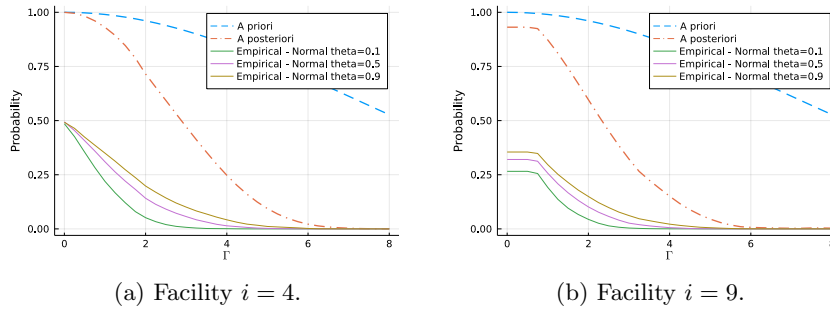
(a) Facility $i = 4$.   (b) Facility $i = 9$.

FIG. 9. *Comparison of a priori and a posteriori bounds when $\tilde{\mathbf{z}}$ admits dependent coordinates. We consider an ellipsoidal uncertainty set, various budgets of uncertainty Γ, and two production capacity constraints.*

7. 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 it 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 [9, 42], 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 approximation of chance constrained conic or matrix inequalities [3, Chapter 10], which has received revived interest recently under the lens of distributionally robust optimization [14, 51, 15].

Appendix A. Preliminary results from probability theory. In this appendix, we recall some useful definitions and results from probability theory regarding sub-Gaussian variables. We refer the reader to [40, 44] for a comprehensive treatment.

DEFINITION A.1 ([40, Definition 1.2]). *A random variable $\tilde{x} \in \mathbb{R}$ is said to be sub-Gaussian with parameter σ^2, denoted $\tilde{x} \sim subG(\sigma^2)$, if $\mathbb{E}[\tilde{x}] = 0$ and $\forall \ s \in \mathbb{R}$,*

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

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 A.2 ([40, Lemma 1.8]). *Let $\tilde{x}$ be a random variable such that $\mathbb{E}[\tilde{x}] = 0$ and $\tilde{x} \in [a, b]$ almost surely. Then, $\tilde{x} \sim subG(\frac{(b-a)^2}{4})$.*

The parameter σ^2 is usually referred to as a *variance proxy* for any sub-Gaussian random variable $\tilde{x}$ with parameter σ^2 satisfying $\mathrm{Var}(\tilde{z}) \leq \sigma^2$—the equality is tight for Gaussian random variables. 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 [3, section 2.4]. Examples from Table 2.3 in [3] for which $\mu_\pm = 0$ satisfy Definition A.1. This definition can be extended to noncentered random variables as well (see [44, Definition 2.2]) and accounts for examples with $\mu_\pm \neq 0$.

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

PROPOSITION A.3 ([40, Lemma 1.3]). *If* $\tilde{x} \sim subG(\sigma^2) \; \forall \; t > 0,$

$$\mathbb{P}\left(\tilde{x} > t\right) \; \leq \; e^{-\frac{t^2}{2\sigma^2}} \quad and \quad \mathbb{P}\left(\tilde{x} < -t\right) \; \leq \; e^{-\frac{t^2}{2\sigma^2}}.$$

Actually, all these tail bounds can be used as equivalent definitions of sub-Gaussian random variables [44, Theorem 2.6].

Relevant for our analysis is the fact that sums of sub-Gaussian random variables are themselves sub-Gaussian as proved in Proposition A.4. The extension to linear combinations of sub-Gaussian random variables is straightforward after observing that if $\tilde{x} \sim \text{subG}(\sigma^2)$, then $\alpha\tilde{x} \sim \text{subG}(|\alpha|^2\sigma^2)$.

PROPOSITION A.4. *Let* $\tilde{x}_i, i = 1, \ldots, n,$ *be* n *sub-Gaussian random variables with respective parameter* σ_i^2.

(a) *If the* x_i*'s are independent, then* $\sum_i \tilde{x}_i$ *is sub-Gaussian with parameter* $\sum_i \sigma_i^2$.
(b) *In general,* $\sum_i \tilde{x}_i$ *is sub-Gaussian with parameter* $\left(\sum_i \sigma_i\right)^2$.

Proof. (a) The first part of the proposition follows from the definition of sub-Gaussian random variables and a Chernoff bound as proved in [40, Theorem 1.6].

(b) We now prove the second part of the result for $n = 2$. Fix $s \in \mathbb{R}$. For any p, q such that $1/p + 1/q = 1$, Hölder's inequality yields

$$\mathbb{E}\left[e^{s\tilde{x}_1 + s\tilde{x}_2}\right] \leq \mathbb{E}\left[e^{s\tilde{x}_1/p}\right]^p \mathbb{E}\left[e^{s\tilde{x}_1/q}\right]^q \leq \exp\left(\frac{s^2}{2}\left[p\sigma_1^2 + q\sigma_2^2\right]\right).$$

Minimizing the right-hand side with respect to p indicates taking $p = 1 + \sqrt{\sigma_2/\sigma_1}$ and $q = 1 + \sqrt{\sigma_1/\sigma_2}$, leading to the following inequality:

$$\mathbb{E}\left[e^{s\tilde{x}_1 + s\tilde{x}_2}\right] \leq \exp\left(\frac{s^2}{2}\left[\sigma_1 + \sigma_2\right]^2\right).$$

The general result follows by induction on n. $\qquad\square$

Appendix B. Omitted proofs.

B.1. Proof of Theorem 3.10.

Proof. Since

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

we can apply Theorem 3.4 (with $b(\mathbf{x}) = f_\star(\mathbf{v}, \mathbf{x})$) and get

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

By definition of the conjugate,

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

which in turns leads to

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

All in all, we have

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

which concludes the proof.   □

B.2. Proof of Proposition 4.2.

Proof. By a change of variable, we have

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

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

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

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

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

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

REFERENCES

[1] O. BARON, J. MILNER, AND H. NASERALDIN, *Facility location: A robust optimization approach*, Prod. Oper. Manage., 20 (2011), pp. 772–785.

[2] A. BEN-TAL, D. DEN HERTOG, AND J.-P. VIAL, *Deriving robust counterparts of nonlinear uncertain inequalities*, Math. Program., 149 (2015), pp. 265–299.

[3] A. BEN-TAL, L. EL GHAOUI, AND A. NEMIROVSKI, *Robust Optimization*, Princeton Ser. Appl. Math. 28, Princeton University Press, 2009.

[4] A. BEN-TAL AND A. NEMIROVSKI, *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Program., 88 (2000), pp. 411–424.

[5] D. BERTSIMAS, D. B. BROWN, AND C. CARAMANIS, *Theory and applications of robust optimization*, SIAM Rev., 53 (2011), pp. 464–501, https://doi.org/10.1137/080734510.

[6] D. BERTSIMAS, V. GUPTA, AND N. KALLUS, *Data-driven robust optimization*, Math. Program., 167 (2018), pp. 235–292.

[7] D. BERTSIMAS, D. PACHAMANOVA, AND M. SIM, *Robust linear optimization under general norms*, Oper. Res. Lett., 32 (2004), pp. 510–516.

[8] D. BERTSIMAS AND M. SIM, *The price of robustness*, Oper. Res., 52 (2004), pp. 35–53.

[9] D. BERTSIMAS AND M. SIM, *Tractable approximations to robust conic optimization problems*, Math. Program., 107 (2006), pp. 5–36.

[10] C. E. BONFERRONI, *Teoria statistica delle classi e calcolo delle probabilita*, Pubbl. R Ist. Sup. Sci. Econom. Commericiali Firenze, 8 (1936), pp. 3–62.

[11] G. CALAFIORE AND M. C. CAMPI, *Uncertain convex programs: Randomized solutions and confidence levels*, Math. Program., 102 (2005), pp. 25–46.

[12] G. C. CALAFIORE AND M. C. CAMPI, *The scenario approach to robust control design*, IEEE Trans. Automat. Control, 51 (2006), pp. 742–753.

[13] M. C. CAMPI AND S. GARATTI, *The exact feasibility of randomized solutions of uncertain convex programs*, SIAM J. Optim., 19 (2008), pp. 1211–1230, https://doi.org/10.1137/07069821X.

[14] W. CHEN, M. SIM, J. SUN, AND C.-P. TEO, *From CVaR to uncertainty set: Implications in joint chance-constrained optimization*, Oper. Res., 58 (2010), pp. 470–485.

[15] Z. CHEN, D. KUHN, AND W. WIESEMANN, *Data-Driven Chance Constrained Programs over Wasserstein Balls*, preprint, https://arxiv.org/abs/1809.00210, 2018.

[16] D. P. DE FARIAS AND B. VAN ROY, *On constraint sampling in the linear programming approach to approximate dynamic programming*, Math. Oper. Res., 29 (2004), pp. 462–478.

[17] E. DELAGE AND S. MANNOR, *Percentile optimization for Markov decision processes with parameter uncertainty*, Oper. Res., 58 (2010), pp. 203–213.

[18] E. DELAGE AND Y. YE, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Oper. Res., 58 (2010), pp. 595–612.

[19] L. DÜMBGEN AND G. WALTHER, *Rates of convergence for random approximations of convex sets*, Adv. in Appl. Probab., 28 (1996), pp. 384–393.

[20] E. ERDOĞAN AND G. IYENGAR, *Ambiguous chance constrained problems and robust optimization*, Math. Program., 107 (2006), pp. 37–61.

[21] P. M. ESFAHANI AND D. KUHN, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, Math. Program., 171 (2018), pp. 115–166.

[22] V. GABREL, C. MURAT, AND A. THIELE, *Recent advances in robust optimization: An overview*, Eur. J. Oper. Res., 235 (2014), pp. 471–483.

[23] R. GAO AND A. J. KLEYWEGT, *Distributionally Robust Stochastic Optimization with Wasserstein Distance*, preprint, https://arxiv.org/abs/1604.02199, 2016.

[24] J. GOH AND M. SIM, *Distributionally robust optimization and its tractable approximations*, Oper. Res., 58 (2010), pp. 902–917.

[25] V. GUPTA, *Near-optimal Bayesian ambiguity sets for distributionally robust optimization*, Management Sci., 65 (2019), pp. 4242–4260.

[26] Y. A. GUZMAN, L. R. MATTHEWS, AND C. A. FLOUDAS, *New a priori and a posteriori probabilistic bounds for robust counterpart optimization* I: *Unknown probability distributions*, Comput. Chem. Engrg., 84 (2016), pp. 568–598.

[27] Y. A. GUZMAN, L. R. MATTHEWS, AND C. 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*, Comput. Chem. Engrg., 101 (2017), pp. 279–311.

[28] Y. A. GUZMAN, L. R. MATTHEWS, AND C. 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*, Comput. Chem. Engrg., 103 (2017), pp. 116–143.

[29] G. A. HANASUSANTO, V. ROITCH, D. KUHN, AND W. WIESEMANN, *A distributionally robust perspective on uncertainty quantification and chance constrained programming*, Math. Program., 151 (2015), pp. 35–62.

[30] G. A. HANASUSANTO, V. ROITCH, D. KUHN, AND W. WIESEMANN, *Ambiguous joint chance constraints under mean and dispersion information*, Oper. Res., 65 (2017), pp. 751–767.

[31] K. HOLMBERG, M. RÖNNQVIST, AND D. YUAN, *An exact algorithm for the capacitated facility location problems with single sourcing*, Eur. J. Oper. Res., 113 (1999), pp. 544–559.

[32] L. J. HONG, Z. HUANG, AND H. LAM, *Learning-based robust optimization: Procedures and statistical guarantees*, Management Sci., 67 (2021), pp. 3447–3467, https://doi.org/10.1287/mnsc.2020.3640.

[33] T. KANAMORI AND A. TAKEDA, *Worst-case violation of sampled convex programs for optimization with uncertainty*, J. Optim. Theory Appl., 152 (2012), pp. 171–197.

[34] D. KUHN, P. M. ESFAHANI, V. A. NGUYEN, AND S. SHAFIEEZADEH-ABADEH, *Wasserstein distributionally robust optimization: Theory and applications in machine learning*, in Operations Research & Management Science in the Age of Analytics, INFORMS, 2019, pp. 130–166.

[35] Z. LI, Q. TANG, AND C. A. FLOUDAS, *A comparative theoretical and computational study on robust counterpart optimization* II: *Probabilistic guarantees on constraint satisfaction*, Ind. Eng. Chem. Res., 51 (2012), pp. 6769–6788.

[36] J. LUEDTKE AND S. AHMED, *A sample approximation approach for optimization with probabilistic constraints*, SIAM J. Optim., 19 (2008), pp. 674–699, https://doi.org/10.1137/070702928.

[37] A. W. MARSHALL AND I. OLKIN, *Multivariate Chebyshev inequalities*, Ann. Math. Statist., 31

      (1960), pp. 1001–1014.

[38] H. NAMKOONG AND J. C. DUCHI, *Stochastic gradient methods for distributionally robust optimization with f-divergences*, Adv. Neural Inform. Process. Syst., 29 (2016), pp. 2208–2216.

[39] A. NEMIROVSKI AND A. SHAPIRO, *Scenario Approximations of Chance Constraints*, Springer, 2006.

[40] P. RIGOLLET AND J.-C. HÜTTER, *High Dimensional Statistics*, Lecture Notes for Course 18.657, MIT, 2017, http://www-math.mit.edu/~rigollet/PDFs/RigNotes17.pdf.

[41] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, 2015.

[42] E. ROOS, D. DEN HERTOG, A. BEN-TAL, F. DE RUITER, AND J. ZHEN, *Approximation of hard uncertain convex inequalities*, available on Optimization Online, June 2018.

[43] B. P. VAN PARYS, P. M. ESFAHANI, AND D. KUHN, *From data to decisions: Distributionally robust optimization is optimal*, Management Sci., 67 (2021), pp. 3387–3402, https://doi.org/10.1287/mnsc.2020.3678.

[44] M. J. WAINWRIGHT, *High-Dimensional Statistics: A Non-asymptotic Viewpoint*, Cambridge Ser. Statist. Probab. Math. 48, Cambridge University Press, 2019.

[45] W. XIE, *On distributionally robust chance constrained programs with Wasserstein distance*, Math. Program. Ser. A, 186 (2021), pp. 115–155.

[46] Y. XIE, J. SNOEYINK, AND J. XU, *Efficient algorithm for approximating maximum inscribed sphere in high dimensional polytope*, in Proceedings of the Twenty-Second Annual Symposium on Computational Geometry, ACM, 2006, pp. 21–29.

[47] H. XU, C. CARAMANIS, AND S. MANNOR, *Optimization under probabilistic envelope constraints*, Oper. Res., 60 (2012), pp. 682–699.

[48] W. YANG AND H. XU, *Distributionally robust chance constraints for non-linear uncertainties*, Math. Program., 155 (2016), pp. 231–265.

[49] İ. YANIKOĞLU AND D. DEN HERTOG, *Safe approximations of ambiguous chance constraints using historical data*, INFORMS J. Comput., 25 (2013), pp. 666–681.

[50] J. ZHEN AND D. DEN HERTOG, *Computing the maximum volume inscribed ellipsoid of a polytopic projection*, INFORMS J. Comput., 30 (2018), pp. 31–42.

[51] S. ZYMLER, D. KUHN, AND B. RUSTEM, *Distributionally robust joint chance constraints with second-order moment information*, Math. Program., 137 (2013), pp. 167–198.