JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 130, No. 1, pp. 1–22, July 2006 (© 2006)
DOI: 10.1007/s10957-006-9084-x

On Distributionally Robust Chance-Constrained Linear Programs[1]

G. C. CALAFIORE[2] AND L. EL GHAOUI[3]

Communicated by B. T. Polyak

Published Online: 1 December 2006

Abstract. In this paper, we discuss linear programs in which the data that specify the constraints are subject to random uncertainty. A usual approach in this setting is to enforce the constraints up to a given level of probability. We show that, for a wide class of probability distributions (namely, radial distributions) on the data, the probability constraints can be converted explicitly into convex second-order cone constraints; hence, the probability-constrained linear program can be solved exactly with great efficiency. Next, we analyze the situation where the probability distribution of the data is not completely specified, but is only known to belong to a given class of distributions. In this case, we provide explicit convex conditions that guarantee the satisfaction of the probability constraints for any possible distribution belonging to the given class.

Key Words. Chance-constrained optimization, probability-constrained optimization, uncertain linear programs, robustness, convex second-order cone constraints.

1. Introduction

In this paper, we study a class of uncertain linear programs of the from

$$\min_{x\in\mathbb{R}^n} \quad c^T x, \tag{1}$$

$$\text{s.t.} \quad a_i^T x + b_i \leq 0, \qquad i = 1, \ldots, m, \tag{2}$$

where $x \in \mathbb{R}^n$ is the decision variable and the problem data $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \ldots, m$ are uncertain. Specifically, we consider here the situation where

[1] This work was supported by FIRB funds from the Italian Ministry of University and Research.

[2] Associate Professor, Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy.

[3] Associate Professor, Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, California.

the uncertainty in the data is of stochastic nature; i.e., the data vectors $d_i \doteq [a_i^T, b_i]^T$, $i = 1, \ldots, m$, are independent $(n+1)$-dimensional random vectors.

A classical approach (see for instance Refs. 1–2) to the solution of (1)–(2) under random uncertainty is to introduce risk levels $\epsilon_i \in (0, 1)$, $i = 1, \ldots, m$, and to enforce the constraints (2) in probability, thus obtaining a so-called chance-constrained linear program (CCLP, a term apparently coined by Charnes and Cooper in Ref. 3) of the form

$$\min_{x \in \mathbb{R}^n} \quad c^T x, \tag{3}$$

$$\text{s.t.} \quad \text{Prob}\left\{a_i^T x + b_i \leq 0\right\} \geq 1 - \epsilon_i, \quad i = 1, \ldots, m. \tag{4}$$

There exist a significant literature on such kind of problems, most of which is resumed in the thorough account given in Ref. 1; see also Ref. 4 for an up-to-date discussion and several applications. We do not intend here to survey all the relevant stochastic programming literature, but mention briefly some of the fundamental issues. A first problem is to determine under which hypotheses on the distribution of the d_i, the optimization problem (3)–(4) is a convex program. A classical result in this sense was first given in Ref. 5, stating that, if d_i is Gaussian, then the corresponding chance constraint imposes a convex (and indeed conic quadratic) constraint on x. Similarly, it can be shown (see Theorem 10.2.1 in Ref. 1) that, if a_i is fixed (nonrandom) and b_i has a log-concave probability density, then the corresponding chance constraint is also convex. An extension of this result to the case when both a_i, b_i have joint log-concave and symmetric density has been given recently in Ref. 6. Another problem relates to how to convert explicitly the probability constraint into a deterministic one, once the probability density of d_i is assigned. Again, this can be done easily in the case of Gaussian distribution, while seemingly little is available directly in the literature for the case of other distributions.

In this paper, we discuss several new aspects of the chance-constrained linear program (3)–(4). First, we analyze the probability constraints for a class of radially symmetric probability distributions and construct explicitly the deterministic convex counterparts of the chance constraints for any distribution in this family (which includes for instance Gaussian, truncated Gaussian, as well as uniform distributions on ellipsoidal support, and also nonunimodal densities).

Next, we move to the main focus of the paper, which deals with chance constraints under distribution uncertainty. The objective of this study is to obtain deterministic restrictions on the variable x, such that the probability constraint

$$\text{Prob}\{a^T x + b \leq 0\} \geq 1 - \epsilon \tag{5}$$

is guaranteed irrespective of the probability distribution of the data $d = [a^T b]^T$, where we denote now with d the generic data vector d_i, $i = 1, \ldots, m$. In this context, we consider three different situations: In a first scenario, we assume that the distribution of d is unknown, but the first two moments (or the mean and the covariance) of d are known. Therefore, the probability constraint (5) needs to be enforced over all possible distributions compatible with the given moments,

$$\inf_{d\sim(\hat{d},\Gamma)} \mathrm{Prob}\{a^T x + b \leq 0\} \geq 1 - \epsilon, \tag{6}$$

where the inf is taken over all distributions with mean $\hat{d}$ and covariance matrix Γ.

In a second scenario, we consider the common situation where the coefficients of the linear program are known only to lie in independent intervals. The widths of the intervals are known and no other information on the distribution of the parameter inside the intervals is available. As in the previous case, we provide explicit conditions for the enforcement of the probability constraint, robustly with respect to the parameter distribution. In the third situation, we consider distributional robustness within the family of radially-symmetric nonincreasing densities introduced in Refs. 7–8.

Lastly, we explore further problem (6), removing the assumption that the exact mean and covariance of d are known. We assume instead that only a certain number N of independent realization $d^{(1)}, d^{(2)}, \ldots, d^{(N)}$ of the random vector d are available and that the empirical estimates of the mean and covariance matrix have been obtained based on this observed sample. This setup raises several fundamental questions, since the empirical estimates are themselves random quantities and cannot be used directly in place of the exact and unknown mean and covariance. Also in this situation, we provide an explicit counterpart of the distributionally robust chance constraint, using recently developed finite-sample results from statistical learning theory.

1.1. Setup and Notation. Under the standing assumption that the random constraints (4) are independent, without loss of generality, we concentrate on a single generic constraint of the form

$$\mathrm{Prob}\{a^T x + b \leq 0\} \geq 1 - \epsilon, \quad \epsilon \in (0, 1), \tag{7}$$

and define the random vector

$$d \doteq [a^T b]^T \in \mathbb{R}^{n+1}, \qquad \text{with } a \in \mathbb{R}^n, b \in \mathbb{R},$$

and

$$\hat{d}^T \doteq E\{d^T\} = E\{[a^T b]\} \doteq [\hat{a}^T \hat{b}],$$

$$\Gamma \doteq \mathrm{var}\{d\} = \mathrm{var}\{[a^T b]^T\} \doteq \begin{bmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{12}^T & \gamma_{22} \end{bmatrix} \succeq 0.$$

We denote with $\upsilon \leq n+1$ the rank of Γ and with $\Gamma_f \in \mathbb{R}^{n+1,\upsilon}$ a full-rank factor such that $\Gamma = \Gamma_f \Gamma_f^T$. Setting

$$\tilde{x} \doteq [x^T\, 1]^T \in \mathbb{R}^{n+1},$$

we define further the scalar quantity

$$\varphi(x) \doteq d^T \tilde{x},$$

and

$$\hat{\varphi}(x) \doteq E\{\varphi(x)\} = \hat{d}^T \tilde{x}, \tag{8}$$

$$\sigma^2(x) \doteq \operatorname{var}\{\varphi(x)\} = \tilde{x}^T \Gamma \tilde{x}. \tag{9}$$

We define also the normalized random variable

$$\tilde{\varphi}(x) = [\varphi(x) - \hat{\varphi}(x)]/\sigma(x),$$

which has zero mean and unit variance. With this notation, the constraint (7) is equivalently rewritten as

$$\operatorname{Prob}\{\varphi(x) \leq 0\} = \operatorname{Prob}\{\tilde{\varphi}(x) \leq -\hat{\varphi}(x)/\sigma(x)\} \geq 1 - \epsilon.$$

In the sequel, $\|x\|$ denotes the Euclidean norm of the vector x, while $\|x\|_\infty$ and $\|x\|_1$ denote the ℓ_∞ and ℓ_1 norms respectively; i.e.,

$$\|x\|_\infty = \max_i |x_i|, \qquad \|x\|_1 = \sum_i |x_i|.$$

For matrices, $\|X\|_F$ denotes the Frobenius norm,

$$\|X\|_F = \sqrt{\operatorname{trace} X\, X^T}.$$

The notation $X \succ 0$ [resp. $X \succeq 0$] is used to denote a symmetric positive-definite [resp. positive - semidefinite] matrix. I_n denotes the identity matrix of dimension n; the subscript n is omitted when it may be inferred easily from context.

2. Chance Constraints under Radial Distributions

In this section, we show that, for a significant class of probability distributions on d, the chance constraint (7) can be expressed explicitly as a deterministic convex constraint on x. We introduce next the class of multivariate distributions of interest.

Definition 2.1. A random vector $d \in \mathbb{R}^{n+1}$ has a Q-radial distribution with defining function $g(\cdot)$ if $d - E\{d\} = Q\omega$ where $Q \in \mathbb{R}^{n+1,\upsilon}$, $\upsilon \leq n+1$, is a fixed, full-rank matrix and $\omega \in \mathbb{R}^\upsilon$ is a random vector having probability density f_w that depends only on the norm of ω; i.e.,

$$f_\omega(\omega) = g(\|\omega\|). \tag{10}$$

The function $g(\cdot)$ that defines the radial shape of the distribution is named the defining function of d.

Remark 2.1. Note that, in the above definition, the covariance of ω may be computed as

$$\Sigma_\omega = \left(V_\upsilon \int_0^\infty r^{\upsilon+1} g(r) \mathrm{d}r \right) I_\upsilon,$$

where V_υ denotes the volume of the Euclidean ball of unit radius in $\mathbb{R}^\upsilon$ (see e.g. Ref. 9); hence, the covariance of d is

$$\mathrm{var}\{d\} = \left(V_\upsilon \int_0^\infty r^{\upsilon+1} g(r) \mathrm{d}r \right) Q Q^T.$$

To be consistent with our notation, according to which $\mathrm{var}\{d\} \doteq \Gamma$, we shall choose Q as

$$Q = \nu \Gamma_f, \quad \nu \doteq \left(V_\upsilon \int_0^\infty r^{\upsilon+1} g(r) \mathrm{d}r \right)^{-1/2}, \tag{11}$$

where Γ_f is a full-rank factor of Γ (i.e., $\Gamma = \Gamma_f \Gamma_f^T$).

Remark 2.2. According to the above definition, the possibly singular multivariate Gaussian distribution in $\mathbb{R}^{n+1}$, with mean $\hat{d}$ and covariance Γ, is Q-radial with $Q = \Gamma_f$ and defining function

$$g(r) = \left(1/(2\pi)^{\upsilon/2}\right) \exp\left(-r^2/2\right).$$

The family of Q-radial distributions includes all probability densities whose level sets are ellipsoids with shape matrix $Q \succ 0$ and may have any radial behavior. Another notable example is the uniform density over an ellipsoidal set, which is discussed further in Section 2.2.

Now, if d is Q-radial with defining function $g(\cdot)$ and covariance Γ (i.e., Q is given by (11)), we have

$$\bar{\varphi}(x) = \tilde{x}^T Q \omega / \sigma(x) = \nu \tilde{x}^T \Gamma_f \omega / \sqrt{\tilde{x}^T \Gamma \tilde{x}}\ ,$$

with ω distributed according to (10). This means that $\bar{\varphi}(x)$ is obtained compressing the random vector $\omega \in \mathbb{R}^\upsilon$ into one-dimension by means of the above scalar product. It results that the distribution of $\bar{\varphi}(x)$ is symmetric around the origin; see for instance Theorem 1 of Ref. 10 for a proof. A similar result is discussed also in Ref. 11 for the special case of uniform distributions (i.e. $g(\cdot)$ constant),

$$f_{\bar{\varphi}(x)/\nu}(\xi) = S_{\upsilon-1} \int_0^\infty g\left(\sqrt{\rho^2 + \xi^2}\right)\rho^{\upsilon-2}\mathrm{d}\rho, \tag{12}$$

where S_n denotes the surface measure of the Euclidean ball of unit radius in $\mathbb{R}^n$. Notice that the above probability density is independent of x. Next, we write

$$\begin{aligned}\mathrm{Prob}\{\varphi(x) \le 0\} &= \mathrm{Prob}\{\bar{\varphi}(x) \le \hat{\varphi}(x)/\sigma(x)\}\\ &= \mathrm{Prob}\{\bar{\varphi}(x)/\nu \le -\hat{\varphi}(x)/\nu\sigma(x)\}\\ &= \Psi(-\hat{\varphi}(x)/\nu\sigma(x))\end{aligned}$$

where we defined the cumulative probability function

$$\Psi(\zeta) \;=\; \mathrm{Prob}\{\bar{\varphi}(x)/\nu \le \zeta\} = \int_{-\infty}^{\zeta} f_{\bar{\varphi}(x)/\nu}(\xi)\mathrm{d}\xi.$$

Therefore,

$$\mathrm{Prob}\{\varphi(x) \le 0\} \ge 1 - \epsilon$$

holds if and only if

$$\bar{\varphi}(x)/\nu\sigma(x) \ge \Psi^{-1}(1 - \epsilon).$$

Notice that, since $f_{\bar{\varphi}(x)/\nu}$ is symmetric around the origin, then $\Psi^{-1}(1-\epsilon)$ is nonnegative if and only if $\epsilon \le 0.5$. Thus, defining

$$\kappa_{\epsilon,r} = \nu\Psi^{-1}(1 - \epsilon), \quad \text{for } \epsilon \in (0, 0.5],$$

we have that the probability constraint in the generic radial case is equivalent to the explicit deterministic constraint

$$\kappa_{\epsilon,r}\sigma(x) + \hat{\varphi}(x) \le 0.$$

Recalling definitions (8), (9), we conclude that (2) is a convex constraint on x; in particular, it is a second-order cone (SOC) convex constraint (see e.g. Ref. 12). Therefore, we have proved the following theorem.

Theorem 2.1. For any $\epsilon \in (0, 0.5]$, the chance constraint

$$\mathrm{Prob}\{d^T\tilde{x} \le 0\} \ge 1 - \epsilon,$$

where d has a Q-radial distribution with defining function $g(\cdot)$ and covariance Γ, is equivalent to the convex second-order cone constraint

$$\kappa_{\epsilon,r}\sigma(x) + \hat{\varphi}(x) < 0,$$

where

$$\kappa_{\epsilon,r} = \nu\Psi^{-1}(1-\epsilon),$$

with Ψ the cumulative probability function of the density (12) and ν given by (11).

In some cases of interest, the cumulative distribution, and hence the corresponding safety parameter $\kappa_{\epsilon,r}$, can be computed in closed form. For instance, this is the case for the Gaussian and the uniform distribution over ellipsoidal support, which are considered next.

2.1. Gaussian Distribution. We have remarked already that a Gaussian distribution with mean $\hat{d}$ and covariance Γ is Q-radial with $Q = \Gamma_f$, $\nu = 1$, and defining function

$$g(r) = \left[1/(2\pi)^{\upsilon/2}\right]\exp(-r^2/2).$$

Consequently, $f_{\bar{\varphi}(x)/\nu}$ from (12) is the Gaussian density function

$$f_{\bar{\varphi}(x)/\nu}(\xi) = (1/\sqrt{2\pi})\exp\left(-\xi^2/2\right)$$

and Ψ is the standard Gaussian cumulative probability function

$$\Psi(\xi) = \Psi_G(\xi) = (1/\sqrt{2\pi})\int_{-\infty}^{\xi}\exp(-t^2/2)\mathrm{d}t.$$

Therefore, for $\epsilon \in (0, 0.5]$, the safety parameter κ_ϵ is given by

$$\kappa_\epsilon = \kappa_{\epsilon,G} = \Psi_G^{-1}(1-\epsilon).$$

Notice that here we recover a classical result (see Ref. 1, Theorem 10.4.1), that could be derived also more directly, without passing through the radial densities framework.

2.2. Uniform Distribution on Ellipsoidal Support.

Lemma 2.1. Let $d - \hat{d} \in \mathbb{R}^{n+1}$ be uniformly distributed in the ellipsoid

$$\mathcal{E} = \{\xi = Qz : \|z\| \leq 1\},$$

where

$$Q \doteq \nu\Gamma_f, \quad \Gamma \succ 0, \quad \nu \doteq \sqrt{n+3}.$$

Then, for any $\epsilon \in (0, 0.5]$, the chance constraint

$$\mathrm{Prob}\{d^T\tilde{x} \leq 0\} \geq 1-\epsilon$$

is equivalent to the convex second-order cone constraint

$$\kappa_{\epsilon,u}\sigma(x) + \hat{\varphi}(x) \leq 0,$$

where

$$\kappa_{\epsilon,u} = \nu\sqrt{\Psi_{\text{beta}}^{-1}(1-2\epsilon)},$$

with $\Psi_{\text{beta}}(\cdot)$ the cumulative distribution of a beta$(1/2; n/2+1)$ probability density.

Proof. We observe first that the uniform distribution in the ellipsoid

$$\mathcal{E} = \{\xi = \nu\Gamma_f z : \|z\| \leq 1\}$$

is obtained multiplying by $\nu\Gamma_f$ a vector ω which is uniformly distributed in the unit Euclidean ball (nonsingular linear transformations preserve uniformity); therefore, if $d - \hat{d}$ is uniform in $\mathcal{E}$, it can be expressed as

$$d - \hat{d} = \nu\Gamma_f\omega,$$

where $\omega \in \mathbb{R}^{n+1}$ is uniform in $\{z : \|z\| \leq 1\}$; i.e.,

$$f_\omega(\omega) = g(\|\omega\|) = \begin{cases} 1/V_{n+1}, & \text{if } \|\omega\| \leq 1, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

and V_{n+1} denotes the volume[4] of the Euclidean ball of unit radius in $\mathbb{R}^{n+1}$. It follows that d is Q-radial with $Q = \nu\Gamma_f$ and defining function $g(\cdot)$ given in (13). Notice that this specific choice of the parameter ν is made according to (11), in order to fix the covariance of d to be equal to Γ. With this defining function, we can solve explicitly the integral in (12), obtaining the density

$$f_{\varphi(x)/\nu}(\xi) = \frac{S_n}{nV_{n+1}}(1-\xi^2)^{n/2}, \quad \xi \in [-1, 1].$$

Since this density is centrally symmetric, we have that the cumulative probability function $\Psi(\xi)$, $\xi \in [-1, 1]$, is given by

$$\Psi(\xi) = 1/2 + (1/2)\text{sign}(\xi)\int_0^{|\xi|}(2S_n/nV_{n+1})(1-t^2)^{n/2}\text{d}t.$$

With the change of variable $z = t^2$, we finally obtain

$$\Psi(\xi) = (1/2) + (1/2)\text{sign}(\xi)\Psi_{\text{beta}}(\xi^2),$$

where Ψ_{beta} denotes the cumulative distribution of a beta $(1/2; n/2+1)$ probability density (see for instance Ref. 13 for the definition of beta densities). The statement

[4] We recall that $V_n = \pi^{n/2}/\Gamma(n/2+1)$, where $\Gamma(\cdot)$ denotes the standard gamma function, and that $S_n = nV_n$.

of the lemma then follows applying Theorem 2.1, where the safety parameter k_e for $\epsilon \in (0, 05]$ is given by

$$\kappa_\epsilon = \kappa_{\epsilon,u} = \nu\sqrt{\Psi_{\text{beta}}^{-1}(1 - 2\epsilon)}. \qquad \square$$

Remark 2.3. Contrary to the Gaussian case, the safety parameter for the uniform distribution on ellipsoidal support depends on the problem dimension n. It may be verified numerically that, the covariance matrices being equal, the safety parameter $\kappa_{\epsilon,u}$ relative to the uniform distribution is smaller (for $\epsilon < 0.5$) than the one relative to the Gaussian distribution, and tends to this latter one as n increases.

Remark 2.4. It is worth to highlight the close relation that exists between chance constraints and deterministically robust constraints. In particular, in robust optimization (see e.g. Ref. 14), one seeks to satisfy the constraints for all possible values of some uncertain but bounded parameters; in the chance constrained setup, one seeks to satisfy the constraints only with high probability. There exist however a complete equivalence between these two paradigms, at least for the simple ellipsoidal model discussed below.

Consider a linear constraint $d^T\tilde{x} \leq 0$, where the data d lie in the ellipsoid

$$\mathcal{E} = \{\hat{d} + \kappa\Gamma_f z : \|z\| \leq 1\}, \quad \Gamma \succ 0,$$

and suppose that we want to enforce the constraint robustly, i.e., for all $d \in \mathcal{E}$. It is readily seen that the set of x that satisfies

$$d^T\tilde{x} \leq 0, \quad \forall d \in \mathcal{E},$$

is the convex quadratic set

$$\{x : \kappa\sqrt{\tilde{x}^T\Gamma\tilde{x}} + \hat{d}^T\tilde{x} \leq 0\},$$

which in our notation reads

$$\{x : \kappa\sigma(x) + \hat{\varphi}(x) \leq 0\}.$$

From this, it follows that a chance constraint of the kind discussed previously can be interpreted as a deterministic robust constraint, for some suitable value of κ and viceversa.

We shall return to the uniform distribution on ellipsoidal support in Section 3.3, where we discuss its role in the context of distributional robustness with respect to a class of symmetric nonincreasing probability densities.

3. Distributionally Robust Chance Constraints

In this section, we present explicit deterministic counterparts of distributionally robust chance constraints. By distributional robustness,we mean that the probability constraint

$$\mathrm{Prob}\{d^T\tilde{x} \le 0\} \ge 1-\epsilon$$

should be enforced robustly with respect to an entire family $\mathcal{D}$ of probability distributions on the data d; i.e.,we consider the problem of enforcing

$$\inf_{d\sim\mathcal{D}} \mathrm{Prob}\{d^T\tilde{x} \le 0\} \ge 1-\epsilon,$$

where the notation $d \sim \mathcal{D}$ means that the distribution of d belongs to the family $\mathcal{D}$. We discuss in particular three classes of distributions: In Section 3.1, we consider the family of all distributions having given mean and covariance; in Section 3.2, we study the family of generic distributions over independent bounded intervals; finally, in Section 3.3, we consider the family of radially symmetric nonincreasing distributions discussed in Refs. 7–8.

3.1. Distributions with Known Mean and Covariance. The first problem that we consider is one where the family $\mathcal{D}$ is composed of all distributions having given mean $\hat{d}$ and covariance Γ. We denote this family with $\mathcal{D} = (\hat{d}, \Gamma)$. The following theorem holds.

Theorem 3.1. For any $\epsilon \in (0, 1)$, the distributionally robust chance constraint

$$\inf_{d\sim(\hat{d},\Gamma)} \mathrm{Prob}\{d^T\tilde{x} \le 0\} \ge 1-\epsilon \tag{14}$$

is equivalent to the convex second-order cone constraint

$$\kappa_\epsilon \sigma(x) + \hat{\varphi}(x) \le 0, \quad \kappa_\epsilon = \sqrt{(1-\epsilon)/\epsilon}. \tag{15}$$

Proof. We express first d as

$$d = \hat{d} + \Gamma_f z,$$

where

$$E\{z\} = 0, \quad \mathrm{var}\{z\} = I,$$

and consider initially the case (a) when $\Gamma_f^T\tilde{x} \ne 0$. Then, from a classical result of Marshall and Olkin (Ref. 15; see also Ref. 16, Theorem 9), we have that

$$\begin{aligned}\sup_{d\sim(\hat{d},\Gamma)} \mathrm{Prob}\{d^T\tilde{x} > 0\} &= \sup_{z\sim(0,I)} \mathrm{Prob}\left\{z^T\Gamma_f^T\tilde{x} > -\hat{d}^T\tilde{x}\right\}\\ &= 1/(1+q^2),\end{aligned}$$

where

$$q^2 = \inf_{z^T \Gamma_f^T \tilde{x} > -\hat{d}^T \tilde{x}} \|z\|^2.$$

We determine a closed-form expression for q^2 as follows. First, we notice that, if $\hat{d}^T \tilde{x} > 0$, then we can just take $z = 0$ and obtain the infimum $q^2 = 0$. Assume then $\hat{d}^T \tilde{x} \leq 0$. Then, the problem amounts to determining the squared distance from the origin of the hyperplane

$$\{z : z^T \Gamma_f^T \tilde{x} = -\hat{d}^T \tilde{x}\},$$

which results to be

$$q^2 = (\hat{d}^T \tilde{x})^2 / \tilde{x}^T \Gamma \tilde{x}.$$

Summarizing, we have

$$q^2 = \begin{cases} 0, & \text{if } \tilde{x}^T \hat{d} = \hat{\varphi}(x) > 0, \\ \hat{\varphi}^2(x)/\sigma^2(x) & \text{if } \hat{\varphi}(x) \leq 0; \end{cases}$$

hence, the constraint (14) is satisfied if and only if

$$1/(1 + q^2) \leq \epsilon,$$

i.e., if and only if

$$\hat{\varphi}(x) \leq 0, \quad \hat{\varphi}^2(x) \geq \sigma^2(x)(1 - \epsilon)/\epsilon,$$

or equivalently if and only if

$$\kappa_\epsilon \sigma(x) \leq -\hat{\varphi}(x), \quad \kappa_\epsilon = \sqrt{(1 - \epsilon)/\epsilon},$$

which proves that, in case (a), (14) is equivalent to (15). On the other hand, in case (b), when $\Gamma_f^T \tilde{x} = 0$, we simply have that

$$\inf_{d \sim (\tilde{d}), \Gamma} \operatorname{Prob}\{d^T \tilde{x} \leq 0\} = 1, \quad \text{if } \hat{\varphi}(x) \leq 0,$$

and it is zero otherwise. Therefore, since $\sigma(x) = 0$, it follows that (14) is still equivalent to (15), which concludes the proof. □

A result analogous to Theorem 3.1 has been employed recently in Ref. 17 in the specific context of kernel methods for classification. An alternative proof based on Lagrangian duality may also be found in Ref. 18. We discuss next the case when additional symmetry information on the distribution is available.

3.1.1. Special Case of Symmetric Distributions. Consider the situation where, in addition to the first moment and covariance of d, we know that the distribution of d is symmetric around the mean. We say that d is symmetric around its mean $\hat{d}$, when $d - \hat{d}$ is symmetric around zero (centrally symmetric), where we define central symmetry as follows.

Definition 3.1. A random vector $\xi \in \mathbb{R}^n$ is centrally symmetric if its distribution μ is such that $\mu(A) = \mu(-A)$, for all Borel sets $A \subseteq \mathbb{R}^n$.

Let $\mathcal{D} = (\hat{d}, \Gamma)_{\mathcal{S}}$ denote the family of symmetric distributions having mean $\hat{d}$ and covariance Γ. The following lemma gives an explicit condition for the satisfaction of the chance constraint robustly over the family $\mathcal{D} = (\hat{d}, \Gamma)_{\mathcal{S}}$.

Lemma 3.1. For any $\epsilon \in (0, 0.5]$, the symmetric distributionally robust chance constraint

$$\inf_{d \sim (\hat{d}, \Gamma)_{\mathcal{S}}} \text{Prob}\{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \tag{16}$$

holds if

$$\kappa_\epsilon \sigma(x) + \hat{\varphi}(x) \leq 0, \quad \kappa_\epsilon = 1/(2\epsilon).$$

Proof. If d is symmetric around $\hat{d}$, then $\varphi(x)$ is symmetric around $\hat{\varphi}(x)$ for all x. Therefore, (16) is satisfied if

$$\sup_{\varphi(x) \sim (\hat{\varphi}(x), \sigma^2(x))_{\mathcal{S}}} \text{Prob}\{\varphi(x) > 0\} \leq \epsilon.$$

We apply now an extension of the Chebychev mean-variance inequality for symmetric distributions (see for instance Ref. 19, Proposition 8): For a random variable z with mean $E\{z\}$, variance $\text{var}\{z\}$, and symmetric distribution, it holds that

$$\sup_{z \sim (E\{z\}, \text{var}\{z\})_{\mathcal{S}}} \text{Prob}\{z > a\} = \begin{cases} (1/2) \min\{1, \text{var}\{z\}/(E\{z\} - a)^2\}, & \text{if } a < E\{z\}, \\ 1, & \text{otherwise.} \end{cases}$$

Hence, in our case, we have that, for $\epsilon \in (0, 0.5]$,

$$\sup_{\varphi(x) \sim (\hat{\varphi}(x), \sigma^2(x))_{\mathcal{S}}} \text{Prob}\{\varphi(x) > 0\} \leq \epsilon$$

is satisfied whenever

$$\sigma^2(x) \leq 2\epsilon \hat{\varphi}(x) \quad \text{and} \quad \hat{\varphi}(x) < 0,$$

i.e., when

$$\sqrt{1/(2\epsilon)} \sigma(x) + \hat{\varphi}(x) \leq 0,$$

from which the statement follows immediately. $\square$

3.2. Random Data in Independent Intervals. In this section, we analyze a data uncertainty model where the random data d have known mean $\hat{d}$ and the individual elements are only known to belong with probability one to independent bounded intervals; i.e., we assume that

$$d_i = \hat{d}_i + \omega_i, \quad i = 1, \dots, n+1,$$

where $\omega \in \mathbb{R}^{n+1}$ is a zero-mean random vector composed of independent elements which are bounded in intervals: $\omega_i \in [\ell_i^-, \ell_i^+]$, $\ell_i^+ \geq 0 \geq \ell_i^-$, $i = 1, \dots, n+1$. Let us denote with $(\hat{d}, L)_{\mathcal{I}}$ the family of distributions on the $(n+1)$-dimensional random variable d satisfying the above condition, where L is a diagonal matrix containing the interval widths,

$$L \doteq \operatorname{diag}(\ell_1^+ - \ell_1^-, \dots, \ell_{n+1}^+ - \ell_{n+1}^-).$$

A first result in this context is stated in the following lemma.

Lemma 3.2. For any $\epsilon \in (0, 1)$, the distributionally robust chance constraint

$$\inf_{d \sim (\hat{d}, L)_{\mathcal{I}}} \operatorname{Prob}\left\{d^T \tilde{x} \leq 0\right\} \geq 1 - \epsilon$$

holds if

$$\sqrt{(1/2)\ln(1/\epsilon)}\,\|L\tilde{x}\| + \hat{\varphi}(x) \leq 0.$$

Proof. By definition, any random vector d whose density belongs to the class $(\hat{d}, L)_{\mathcal{I}}$ is expressed as

$$d_i = \hat{d}_i + \omega_i;$$

therefore,

$$d^T \tilde{x} = \hat{d}^T \tilde{x} + \sum_{i=1}^{n+1} \xi_i,$$

where we define

$$\xi_i \doteq x_i \omega_i, \quad i = 1, \dots, n, \quad \xi_{n+1} \doteq \omega_{n+1}.$$

With our usual notation,

$$\varphi(x) = d^T \tilde{x}, \quad \hat{\varphi}(x) = \hat{d}^T \tilde{x};$$

hence, we have that

$$\operatorname{Prob}\{\varphi(x) \leq 0\} = \operatorname{Prob}\left\{\sum_{i=1}^{n+1} \xi_i \leq -\hat{\varphi}(x)\right\}. \tag{17}$$

Now, the ξ_i are zero-mean, independent, and bounded in intervals of width $|x_i|(\ell_i^+ - \ell_i^-)$, for $i = 1, \dots, n$, and $\ell_{n+1}^+ - \ell_{n+1}^-$ respectively. Therefore, applying the Hoeffding tail probability inequality to (17), see Ref. 20, we obtain that, if $\hat{\varphi}(x) \leq 0$, then

$$\mathrm{Prob}\{\varphi(x) \leq 0\} \geq 1 - \exp\left[\frac{-2\hat{\varphi}^2(x)}{\left(\ell_{n+1}^+ - \ell_{n+1}^-\right)^2 + \sum_{i=1}^n x_i^2\left(\ell_i^+ - \ell_i^-\right)^2}\right],$$

from which the statement follows easily. □

3.3. Radially Symmetric Nonincreasing Distributions. In this section, we consider two classes of radially symmetric nonincreasing distributions (RSNID) whose supports are defined by means of the Euclidean and infinity norms. These are special cases of the RSNIDs introduced in Refs. 7–8, whose support were generic star-shaped sets. Specifically, we define our two distribution classes in a way similar to Definition 2.1. Let us introduce the sets

$$\mathcal{H}(\hat{d}, P) \doteq \{d = \hat{d} + P\omega : \|\omega\|_\infty \leq 1\},$$
$$\mathcal{E}(\hat{d}, Q) \doteq \{d = \hat{d} + Q\omega : \|\omega\| \leq 1\},$$

where

$$P = \mathrm{diag}(p_1, \dots, p_{n+1}) \succ 0, \quad Q \succ 0.$$

Clearly, $\mathcal{H}(\hat{d}, P)$ is an orthotope centered in $\hat{d}$, with half-side lengths specified by P, while $\mathcal{E}(\hat{d}, Q)$ is an ellipsoid centered in $\hat{d}$, with shape matrix Q. The classes of interest are defined as follows.

Definition 3.2. A random vector $d \in \mathbb{R}^{n+1}$ has a probability distribution within the class $\mathcal{F}_\mathcal{H}$ [resp. $\mathcal{F}_\mathcal{E}$] if $d - E\{d\} = P_\omega$ [resp. $d - E\{d\} = Q_\omega$], where ω is a random vector having probability density f_ω such that

$$f_\omega = \begin{cases} g(\|\omega\|_\infty), & \text{for } \|\omega\|_\infty \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$
$$\left[\text{resp.}\, f_\omega(\omega) = \begin{cases} g(\|\omega\|), & \text{for } \|\omega\| \leq 1, \\ 0, & \text{otherwise,} \end{cases}\right]$$

and where $g(\cdot)$ is a nonincreasing function.

Notice that the uniform distribution on ellipsoidal support discussed in Section 2.2 belongs to the class $\mathcal{F}_\mathcal{E}$, while the uniform distribution on the orthotope $\mathcal{H}(\hat{d}, P)$ belongs to the class $\mathcal{F}_\mathcal{H}$. The following proposition, based on the uniformity principle (see Ref. 7), states that these uniform distributions are actually the worst-case distributions in the given classes.

Proposition 3.1. For any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint

$$\inf_{d \sim \mathcal{F}_{\mathcal{H}}} \mathrm{Prob}\{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \tag{18}$$

is equivalent to the chance constraint

$$\mathrm{Prob}\{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad d \sim U(\mathcal{H}(\hat{d}, P)), \tag{19}$$

where $U(\mathcal{H}(\hat{d}, P))$ is the uniform distribution over $\mathcal{H}(\hat{d}, P)$.

Similarly, for any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint

$$\inf_{d \sim \mathcal{F}_{\mathcal{E}}} \mathrm{Prob}\{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon$$

is equivalent to the chance constraint

$$\mathrm{Prob}\{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad d \sim U(\mathcal{E}(\hat{d}, Q)), \tag{20}$$

where $U(\mathcal{E}(\hat{d}, Q))$ is the uniform distribution over $\mathcal{E}(\hat{d}, Q)$.

The proof of this proposition is established readily from Theorem 6.3 of Ref. 8. Indeed, it suffices to show (we consider here only the family $\mathcal{F}_{\mathcal{E}}$, the case of the family $\mathcal{F}_{\mathcal{H}}$ being identical) that, under the stated hypotheses, the set

$$\Omega \doteq \{\omega : \hat{d}^T \tilde{x} + \tilde{x}^T Q \omega \leq 0\}$$

is star-shaped, which simply means that $0 \in \Omega$ and that $\omega \in \Omega$, $\rho \in [0, 1]$ implies $\rho\omega \in \Omega$. To show this, first notice that, by central symmetry, for any distribution in $\mathcal{F}_{\mathcal{E}}$,

$$\mathrm{Prob}\{d^T \tilde{x} \leq 0\} \geq 0.5, \quad \text{if and only if } \hat{d}^T \tilde{x} \leq 0.$$

Hence, $\epsilon \in (0, 0.5]$ requires $\hat{d}^T \tilde{x} \leq 0$ and in this situation, $0 \in \Omega$. Consider now any $\omega \in \Omega$ and take $\rho \in [0, 1]$. We have that

$$\tilde{x}^T Q \omega \leq -\hat{d}^T \tilde{x} \Rightarrow \rho \tilde{x}^T Q \omega \leq \rho(-\hat{d}^T \tilde{x});$$

since $(-\hat{d}^T \tilde{x}) \geq 0$ and $\rho \leq 1$, it follows that

$$\rho \tilde{x}^T Q \omega \leq -\hat{d}^T \tilde{x},$$

i.e., $\rho\omega \in \Omega$, proving that Ω is star-shaped.

Remark 3.1. As a result of Proposition 3.1, we have that a distributionally robust constraint over the family $\mathcal{F}_{\mathcal{E}}$ is equivalent to a probability constraint (20) involving the uniform density over ellipsoidal support, which in turn is converted into an explicit second-order cone constraint, using Lemma 2.1.

A distributionally robust constraint over the family $\mathcal{F}_{\mathcal{H}}$ is instead equivalent to a probability constraint (19) involving the uniform density over the orthotope

$\mathcal{H}(\hat{d}, P)$. In Lemma 3.3 below, we provide a new explicit sufficient condition for (19) to hold. In this respect, we notice that it was shown already in Ref. 21 that, for $\epsilon \in (0, 0.5]$,

$$\mathrm{Prob}_{d\sim U(\mathcal{H}(\hat{d}P))}\{d^T\tilde{x} \leq 0\}$$

is a convex constraint on x. However, no explicit condition was given in this Ref. 21 to enforce such convex constraint. In this regard, we remark that the many available convexity results for chance constraints, while interesting in theory, are not very useful in practice unless the chance constraint is converted explicitly into a standard deterministic convex constraint that can be fed to a numerical optimization code. For this reason, in all cases when an explicit description of the chance constraint is too difficult, it appears to be very useful to obtain sufficient conditions that enforce the probability constraint.

Lemma 3.3. For any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint (18) holds if

$$\sqrt{(1/6)\log(1/\epsilon)}\|2P\tilde{x}\| + \hat{\varphi}(x) \leq 0.$$

Proof. We start by establishing a simple auxiliary result. Let ξ be a zero-mean random variable uniformly distributed in the interval $[-c, c]$, $c \geq 0$. Then, for any $\lambda \geq 0$, it holds that

$$\log E\{e^{\lambda\xi}\} \leq \lambda^2 c^2/6. \tag{21}$$

The above fact is proved as follows. Compute in closed form

$$E\{e^{\lambda\xi}\} = \sinh(\lambda c)/\lambda c$$

and consider the function

$$\psi(z) \doteq \log[\sinh(z)/z], \quad z \doteq \lambda c,$$

extended by continuity to $\psi(0) = 0$. Then, we can check by direct calculation that $\psi'(0) = 0$ and $\psi''(0) = 1/3$, where ψ', ψ'' denote respectively the first and second derivatives of ψ and moreover $\psi''(z) \leq 1/3$, $\forall z$. Therefore, by the Taylor expansion with Lagrange remainder, we have that, for some $\theta \in [0, z]$,

$$\psi(z) = \psi(0) + z\psi'(0) + (1/2)z^2\psi''(\theta) \leq (1/6)z^2,$$

from which (21) follows immediately.

Now, we write $d = \hat{d} + P_\omega$ [we recall that $P = \mathrm{diag}(p_1, \ldots, p_{n+1}) \succ 0$] and follow the same steps of the proof of Lemma 3.2 up to (17). Then, we observe that, by Proposition 3.1, the infimum of the probability is attained when the ω_i's are uniformly distributed in $[-1, 1]$. Therefore, in the worst-case, the

ξ_i appearing in (17) are zero-mean, independent, and uniformly distributed in intervals $|x_i|[-p_i, p_i]$, for $i = 1, \dots, n$, and $[-p_{n+1}, p_{n+1}]$ respectively.

By the Chernoff bounding method applied to the Markov probability inequality, we next have that, for $\hat{\varphi}(x) \le 0$ and any $\lambda \ge 0$,

$$\begin{aligned}\text{Prob}\left\{\sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x)\right\} &\le E\left\{e^{\lambda}\sum_{i=1}^{n+1}\xi_i\right\} \Big/ e^{-\lambda\hat{\varphi}(x)}, \\ &= \prod_{i=1}^{n+1} E\left\{c^{\lambda\xi_i}\right\} / e^{-\lambda\hat{\varphi}(x)}.\end{aligned}$$

By (21), we further have

$$E\{e^{\lambda\xi_i}\} \le \mathrm{e}^{(\lambda p_i x_i)^2/6}, \quad i = 1, \dots, n,$$
$$E\{e^{\lambda\xi_{n+1}}\} \le \mathrm{e}^{(\lambda p_{n+1})^2/6}$$

and hence

$$\text{Prob}\left\{\sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x)\right\} \le \mathrm{e}^{\lambda^2\|2P\tilde{x}\|^2/24+\lambda\hat{\varphi}(x)} \le \mathrm{e}^{-6\hat{\varphi}^2(x)/\|2P\tilde{x}\|^2}. \tag{22}$$

where the last inequality obtains selecting $\lambda \ge 0$ so to minimize the bound, which results in

$$\lambda = -12\hat{\varphi}(x)/\|2P\tilde{x}\|^2.$$

Finally, the probability on the LHS of (22) is smaller than $\epsilon \in (0, 0.5]$ if

$$\hat{\varphi}(x) \le 0 \quad \text{and} \quad \|2P\tilde{x}\|^2 \log 1/\epsilon \le 6\hat{\varphi}^2(x),$$

which can be rewritten compactly as the convex second-order cone constraint

$$\sqrt{(1/6)\log(1/\epsilon)}\|2P\tilde{x}\| + \hat{\varphi}(x) \le 0,$$

thus proving the claim. □

Remark 3.2. Notice that the result in Lemma 3.3 improves (decreases) by a factor $\sqrt{3}$ the safety constant with respect to the result given in Lemma 3.2. In fact, the widths of the intervals being equal (i.e. $L = 2P$), the distribution class $(\hat{d}, L)_{\mathcal{I}}$ of Lemma 3.2 includes nonsymmetric and possibly increasing densities; hence, it is richer than the class $\mathcal{F}_{\mathcal{H}}$ of Lemma 3.3.

4. Robustness to Estimation Uncertainty

In Section 3.1, we concentrated our attention on the solution of the chance-constrained problem, assuming that the mean and covariance of the data d

were exactly known. However, In many situations, these quantities need actually to be estimated from empirical data. If a batch of independent extractions $d^1, \ldots, d^N$ from an unknown distribution is available, then the following standard empirical estimates of the true mean $\hat{d}$ and covariance Γ can be formed:

$$\hat{d}_N = (1/N)\sum_{i=1}^{N} d^i,$$

$$\Gamma_N = (1/N)\sum_{i=1}^{N}(d^i - \hat{d}_N)(d^i - \hat{d}_N)^T.$$

Clearly, care should be exerted now, since mere substitution of these estimated values in place of the unknown true ones in the problems discussed previously would not necessarily enforce the correct chance constraints. In the sequel, we present a rigorous approach for taking moment estimation errors into account in the chance constraints. Instrumental to our developments are the following key and surprisingly recent results from Ref. 22 on finite-sample estimation of the mean and covariance matrix. For the mean case, the first lemma below provides for vectors a result similar in spirit to the Hoeffding inequality for scalar random variables.

Lemma 4.1. (See Theorem 3 of Ref. 22.) Let $d^1, \ldots, d^N \in \mathbb{R}^{n+1}$ be extracted independently according to the unknown distribution II and let

$$R = \sup_{d \in \mathrm{support}(II)} \|d\|, \qquad \beta \in (0, 1).$$

Then, with probability at least $1 - \beta$, it holds that

$$\|\hat{d}_N - \hat{d}\| \leq (R/\sqrt{N})\left[2 + \sqrt{2\log(1/\beta)}\right].$$

The next lemma provides a similar concentration inequality for the sample covariance matrix.

Lemma 4.2. (See Corollary 2 of Ref. 22.) Let $d^1, \ldots, d^N \in \mathbb{R}^{n+1}$ be extracted independently according to the unknown distribution II and let

$$R = \sup_{d \in \mathrm{support(II)}} \|d\|, \quad \beta \in (0, 1).$$

Then, provided that

$$N \geq \left[2 + \sqrt{2\log(2/\beta)}\right]^2,$$

it holds with probability at least $1 - \beta$ that

$$\|\Gamma_N - \Gamma\|_F \leq \left[2R^2/\sqrt{N}\right)\left(2 + \sqrt{2\log(2/\beta)}\right].$$

We present next the key result of this section, which specifies how to enforce the chance constraint in the case when only empirical estimates of the mean and covariance are available. It should be clear that, since the estimates $\hat{d}_N, \Gamma_N$ are themselves random variables, the resulting chance constraint cannot be specified deterministically, but will be enforced only up to a given (high) level of probability.

Theorem 4.1. Let $d^1, \ldots, d^N$ be N independent samples of the random vector $d \in \mathbb{R}^{n+1}$ having an unknown distribution II and let $R = \sup_{\|d \in \text{support}(\Pi)} \|d\|$. Let further, $\hat{d}_N, \Gamma_N$ be the sample estimates of the mean and variance of d, computed on the basis of the N available samples; denote with $\hat{d}, \Gamma$ the respective true (unknown) values. Define also

$$r_1 \doteq \left(R/\sqrt{N}\right)\left[2 + \sqrt{2\log(2/\delta)}\right],$$
$$r_2 \doteq \left(2R^2/\sqrt{N}\right)\left[2 + \sqrt{2\log(4/\delta)}\right].$$

Then, for assigned probability levels $\epsilon, \delta \in (0, 1)$, the distributionally robust chance constraint

$$\inf_{d\sim(\hat{d},\Gamma)} \text{Prob}\{d^T\tilde{x} \leq 0\} \geq 1 - \epsilon \tag{23}$$

holds with probability at least $1 - \delta$, provided that

$$N \geq \left[2 + \sqrt{2\log(4/5)}\right]^2,$$
$$\sqrt{(1-\epsilon)/\epsilon}\sqrt{\tilde{x}^T(\Gamma_N + r_2 I)\tilde{x}} + \tilde{d}_N^T\tilde{x} + \|\tilde{x}\| r_1 \leq 0.$$

Proof. Applying Lemma 4.1 with $\beta = \delta/2$, we have that, with probability at least $1 - \delta/2$, it holds that

$$\hat{d} = \hat{d}_N + \xi, \quad \text{for some } \xi \in \mathbb{R}^{n+1} : \|\xi\| \leq r_1. \tag{24}$$

Similarly, applying Lemma 4.2 with $\beta = \delta/2$, we have that, with probability at least $1 - \delta/2$, it holds that

$$\Gamma = \Gamma_N + \Delta, \quad \text{for some } \Delta \in \mathbb{R}^{n+1,n+1} : \|\Delta\|_F \leq r_2, \tag{25}$$

provided that

$$N \geq \left[2 + \sqrt{2\log(4/\delta)}\right]^2.$$

Combining the two events above, we have that equations (24), (25) jointly hold with probability at least

$$(1 - \delta/2)^2 \geq 1 - \delta.$$

Now, from Theorem 3.1, we know that

$$\kappa_\epsilon \sqrt{\tilde{x}^T \Gamma \tilde{x}} + \tilde{d}^T \tilde{x} \le 0, \quad \kappa_\epsilon = \sqrt{(1-\epsilon)/\epsilon} \tag{26}$$

imply the satisfaction of the chance constraint (23). Hence, substituting (24), (25) in (26), we have that (23) holds with probability at least $1-\delta$, provided that the following inequality holds:

$$\kappa_\epsilon \sqrt{\tilde{x}^T (\Gamma_N + \Delta) \tilde{x}} + (\hat{d}_N + \xi)^T \tilde{x} \le 0,$$

for all $\xi : \|\xi\| \le r_1$ and all $\Delta : \|\Delta\|_F \le r_2$. The statement of the theorem then follows from the majorization below

$$\begin{aligned}
&\kappa_\epsilon \sqrt{\tilde{x}^T (\Gamma_N + \Delta) \tilde{x}} + (\hat{d}_N + \xi)^T \tilde{x} \\
&= \kappa_\epsilon \sqrt{\tilde{x}^T \Gamma_N \tilde{x} + \mathrm{trace}(\Delta \tilde{x} \tilde{x}^T)} + (\hat{d}_N + \xi)^T \tilde{x} \\
&\le \kappa_\epsilon \sqrt{\tilde{x}^T \Gamma_N \tilde{x} + \|\Delta\|_F \|\tilde{x}\tilde{x}^T\|_F} + \hat{d}_N^T \tilde{x} + \|\tilde{x}\| \|\xi\| \\
&\le \kappa_\epsilon \sqrt{\tilde{x}^T (\Gamma_N + r_2 I) \tilde{x}} + \hat{d}_N^T \tilde{x} + \|\tilde{x}\| r_1. \qquad \square
\end{aligned}$$

5. Conclusions

In this paper, we discussed several issues related to probability-constrained linear programs. In particular, we provided closed-form expressions for the probability constraints when the data distribution is of radial type (Theorem 2.1). We analzed further in Section 3 the case when the information on the distribution is incomplete and provided explicit results for enforcement of the chance constraints, robustly with respect to the distribution. As discussed in Section 3.3, the uniform density distribution over orthotopes or ellipsoids plays an important role in distributional robustness, since it results to be the worst-case distribution in certain symmetric and non increasing density distribution classes. The results of Section 4 treat instead the case when the mean and covariance of the data distribution are unknown, but can be estimated from available observations. In this respect, Theorem 4.1 provides a convex condition that guarantees (up to a given level of confidence) the satisfaction of the chance constraint for any distribution that could have generated the observed data.

For space reasons, we did not present applications or numerical examples in this paper. However, the interested reader may find two examples of application to optimal portfolio selection and model predictive control problems in the full version of this paper, available in Ref. 23. Other examples of the use of chance constraints in control design have appeared recently in Ref. 6, while Ref. 17 applies a related technique in the context of kernel methods for classification.

References

1. PREKOPA, A., *Stochastic Programming*, Kluwer Academic Publishers, Dodrecht, Netherlands, 1995.
2. VAJDA, S., *Probabilistic Programming*, Academic Press, New York, NY, 1972.
3. CHARNES, A., and COOPER, W., *Deterministic Equivalents for Optimizing and Satisfying under Chance Constraints*, Operations Research, Vol. 11, pp. 18–39, 1963.
4. URYASEV, S. U., Editor, *Probabilistic Constrained Optimization*, Kluwer Academic Publishers, Dodrecht, Netherlands, 2000.
5. VAN DE PANNE, C., and POPP, W., *Minimum Cost Cattle Feed under Probabilistic Problem Constraint*, Management Science, Vol. 9, pp. 405–430, 1963.
6. LAGOA, C., LI, X., and SZNAIER, M., *Application of Probabilistically Constrained Linear Programs to Risk-Adjusted Controller Design*, Proceeding of the American Control Conference, Arlington, Virginia, pp. 738–743, 2001.
7. BARMISH, B., and LAGOA, C., *The Uniform Distribution: A Rigorous Justification for Its Use in Robustness Analysis*, Mathematics of Control, Signals, and Systems, Vol. 10, pp. 203–222, 1997.
8. BARMISH, B., LAGOA, C., and TEMPO, R., *Radially Truncated Uniform Distributions for Probabilistic Robustness of Control Systems*, Proceedings of the American Control Conference, Albuquerque, New Mexico, pp. 853–857, 1997.
9. GUPTA, A., and NAGAR, D., *Matrix Variate Distributions*, Monographs and Surveys in Pure and Applied Mathematics., Chapman and Hall-CRC, Boca Raton, Florida, 1999.
10. CALAFIORE, G., and DABBENE, F., *Loop Gain under Random Feedback*, Proceedings of the IEEE Conference on Decision and Control, Orlando, Florida, 2001.
11. POLYAK, B., and SHCHERBAKOV, P., *Random Spherical Uncertainty in Estimation and Robustness*, IEEE Transactions on Automatic Control, Vol. 45, pp. 2145–2150, 2000.
12. LOBO, M., VANDENBERGHE, L., BOYD, S., and LEBRET, H., *Applications of Second-Order Cone Programming*, Linear Algebra and Its Applications, Vol. 284, pp. 193–228, 1998.
13. PAPOULIS, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, 1991.
14. BEN-TAL, A., and NEMIROVSKI, A., *Robust Solutions of Uncertain Linear Programs*, Operations Research Letters, Vol. 25, pp. 1–13, 1999.
15. MARSHALL, A., and OLKIN, I., *Multivariate Chebyshev Inequalities*, Annals of Mathematical Statistics, Vol. 31, pp. 1001–1014, 1960.
16. BERTSIMAS, D., and POPESCU, I., *Optimal Inequalities in Probability Theory: A Convex Optimization Approach*, Technical Report, Department of Mathematics and Operations Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1998.
17. LANCKRIET, G., EL GHAOUI, L., BHATTACHARYYA, C., and JORDAN, M., *Minimax Probability Machine*, Advances in Neural Information Processing Systems, Vol. 14, 2001.
18. EL GHAOUI, L., OKS, M., and OUSTRY, F., *Worst-Case Value-at-Risk and Robust Portfolio Optimization: A Conic Programming Approach*, Operations Research, Vol. 51, pp. 543–556, 2003.

19. POPESCU, I., *An SDP Approach to Moment Bounds for Distributions with Convex Properties*, Technical Report, INSED Working Paper 2002, available on-line.
20. HOEFFDING, W., *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association, Vol. 58, pp. 13–30, 1963.
21. LAGOA, C., *On the Convexity of Probabilistically Constrained Linear Programs*, Proceedings of the IEEE Conference on Decision and Control, Phoenix, Arizona, pp. 516–521, 1999.
22. SHAWE-TAYLOR, J., and CRISTIANINI, N., *Estimating the Moments of a Random Vector with Applications*, Personal Communication, 2003.
23. CALAFIORE, G., and EL GHAOUI, L., *Distributionally Robust Chance-Constrained Linear Programs with Applications*, Techical Report, DAUIN, Politecnico di Torino, Torino, Italy, 2005 (available at http://staff.polito.it/giuseppe.calafiore/papers/JOTA_final.pdf).