

Distributionally robust chance-constrained linear programs with applications

Giuseppe Calafiore* and Laurent El Ghaoui**

*Dipartimento di Automatica e Informatica
Politecnico di Torino

Corso Duca degli Abruzzi, 24 – 10129 Torino, Italy

** Dept. of Electrical Engineering and Computer Science
University of California at Berkeley, USA

September 5, 2003

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 (i.e. radial distributions) on the data, the probability constraints can be explicitly converted into convex second order cone (SOC) constraints, and hence the probability constrained linear program can be solved exactly with great efficiency. We next analyze the situation when the probability distribution of the data is not completely specified, but it 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. Application examples to portfolio optimization and model predictive control are used to illustrate the results.

1 Introduction

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

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{subject to:} \quad (1.1)$$

$$a_i^T x + b_i \leq 0, \quad i = 1, \dots, m \quad (1.2)$$

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

$$d_i \doteq \begin{bmatrix} a_i \\ b_i \end{bmatrix}, \quad i = 1, \dots, m$$

are independent $(n + 1)$ -dimensional random vectors.

A classical approach (see for instance [20, 22]) to the solution of (1.1)–(1.2) under random uncertainty is to introduce *risk levels* $\epsilon_i \in (0, 1)$, $i = 1, \dots, m$, and to enforce the constraints (1.2) in probability, thus obtaining a so-called ‘chance-constrained’ linear program (CCLP, this term was apparently coined by Charnes and Cooper in [8]) of the form

$$\min_{x \in \mathbb{R}^n} c^T x \quad \text{subject to:} \quad (1.3)$$

$$\text{Prob} \{a_i^T x + b_i \leq 0\} \geq 1 - \epsilon_i, \quad i = 1, \dots, m. \quad (1.4)$$

There exist a significant literature on such kind of problems, most of which is resumed in the thorough account given in [20]; see also [9] for a up-to-date discussion of several applications. We do not intend here to survey all the relevant stochastic programming literature on the topic, but briefly mention some of the fundamental issues. A first problem is to determine under which hypotheses on the distribution of the d_i ’s, the optimization problem (1.3)–(1.4) is a convex program. A classical result in this sense was first given in [23], 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 [20]) that if a_i is fixed (non-random) 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 recently given in [13]. Another problem relates to how to explicitly convert the probability constraint into a deterministic one, once the probability density of d_i is assigned. Again, this can be easily done in the case of Gaussian distribution, while seemingly little is directly available in the literature for the case of other distributions.

In this paper we discuss several new aspects of the chance constrained linear program (1.3)–(1.4). First, we analyze the probability constraints for a class of radially symmetric probability distributions, and *explicitly* construct 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 non-unimodal 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

$$\text{Prob} \{a^T x + b \leq 0\} \geq 1 - \epsilon \quad (1.5)$$

is guaranteed irrespective of the probability distribution of the data $d = [a^T \ b]^T$, where we now denote with d the generic data vector d_i , $i = 1, \dots, m$. 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 (1.5) needs to be enforced over all possible distributions compatible with the given moments:

$$\inf_{d \sim (\hat{d}, \Gamma)} \text{Prob} \{a^T x + b \leq 0\} \geq 1 - \epsilon, \quad (1.6)$$

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 when the coefficients of the linear program are only known 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 non-increasing densities introduced in [1, 2].

Lastly, we further explore problem (1.6), 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 $d^{(1)}, d^{(2)}, \dots, d^{(N)}$ of the random vector d are available, and 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 directly used 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.

To conclude the paper, we present some illustrative examples of application to problems in portfolio optimization and robust control.

1.1 Setup and Notation

Under the standing assumption that the random constraints (1.4) are independent, without loss of generality, we concentrate on a single generic constraint of the form

$$\text{Prob}\{a^T x + b \leq 0\} \geq 1 - \epsilon, \quad (1.7)$$

where $\epsilon \in (0, 1)$, and define the random vector

$$d \doteq \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^{n+1}$$

with $a \in \mathbb{R}^n$, $b \in \mathbb{R}$, and

$$\hat{d} \doteq E\{d^T\} = E\{[a^T \ b]\} \doteq [\hat{a}^T \ \hat{b}]; \quad (1.8)$$

$$\Gamma \doteq \text{var}\{d\} = \text{var}\{[a^T \ b]\} \doteq \begin{bmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{12}^T & \gamma_{22} \end{bmatrix} \succeq 0. \quad (1.9)$$

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

$$\tilde{x} \doteq \begin{bmatrix} x \\ 1 \end{bmatrix} \in \mathbb{R}^{n+1},$$

we further define

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

and

$$\hat{\varphi}(x) \doteq E\{\varphi(x)\} = \hat{d}^T \tilde{x}, \quad (1.10)$$

$$\sigma^2(x) \doteq \text{var}\{\varphi(x)\} = \tilde{x}^T \Gamma \tilde{x}. \quad (1.11)$$

Define also the normalized random variable

$$\bar{\varphi}(x) = \frac{\varphi(x) - \hat{\varphi}(x)}{\sigma(x)} \quad (1.12)$$

which has zero mean and unit variance. With this notation, the constraint (1.7) is equivalently rewritten as

$$\text{Prob}\{\varphi(x) \leq 0\} = \text{Prob}\left\{\bar{\varphi}(x) \leq -\frac{\hat{\varphi}(x)}{\sigma(x)}\right\} \geq 1 - \epsilon. \quad (1.13)$$

In the sequel, $\|x\|$ denotes the Euclidean norm of vector x , while $\|x\|_\infty$ and $\|x\|_1$ denote the ℓ_∞ and ℓ_1 norms, respectively; i.e. $\|x\|_\infty = \max_i |x_i|$, $\|x\|_1 = \sum_i |x_i|$. For matrices, $\|X\|_F$ denotes the Frobenius norm, $\|X\|_F = \sqrt{\text{trace } XX^T}$. The notation $X \succ 0$ (resp. $X \succeq 0$) is used to denote a symmetric positive definite (resp. positive semi-definite) matrix. I_n denotes the identity matrix of dimension n .

2 Chance Constraints Under Radial Distributions

In this section, we show that for a significant class of probability distributions on d , the chance constraint (1.7) can be *explicitly* expressed as a deterministic convex constraint on x . We next introduce 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 $Q \in \mathbb{R}^{n+1,v}$, $v \leq n+1$ is a fixed, full-rank matrix and $\omega \in \mathbb{R}^v$ is a random vector having probability density f_ω that only depends on the norm of ω , i.e.

$$f_\omega(\omega) = g(\|\omega\|). \quad (2.14)$$

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 is

$$\Sigma_\omega = \left(V_v \int_0^\infty r^{v+1} g(r) dr \right) I_v \quad (2.15)$$

where V_v denotes the volume of the Euclidean ball of unit radius in $\mathbb{R}^v$ (see e.g. [10]), and hence the covariance of d is

$$\text{var}\{d\} = \left(V_v \int_0^\infty r^{v+1} g(r) dr \right) QQ^T.$$

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

$$Q = \nu \Gamma_f, \quad \nu \doteq \left(V_v \int_0^\infty r^{v+1} g(r) dr \right)^{-1/2} \quad (2.16)$$

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

$$g(r) = \frac{1}{(2\pi)^{v/2}} \exp(-r^2/2). \quad (2.17)$$

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 further discussed in Section 2.2. $\star$

Now, if d is Q -radial with defining function $g(\cdot)$ and covariance Γ (i.e. Q is given by (2.16)), we have

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

with ω distributed according to (2.14). This means that $\bar{\varphi}(x)$ is obtained ‘compressing’ the random vector $\omega \in \mathbb{R}^v$ into one-dimension, by means of the above scalar product. It results that the distribution of $\bar{\varphi}(x)$ is symmetric around the origin, and in particular (see for instance [7], Theorem 1 for a proof. A similar result for the special case of uniform distributions – i.e. $g(\cdot)$ constant – is also discussed in [18])

$$f_{\bar{\varphi}(x)/\nu}(\xi) = S_{v-1} \int_0^\infty g(\sqrt{\rho^2 + \xi^2}) \rho^{v-2} d\rho, \quad (2.18)$$

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

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

where we defined the cumulative probability function

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

Therefore,

$$\text{Prob} \{ \varphi(x) \leq 0 \} \geq 1 - \epsilon$$

holds if and only if

$$-\frac{\hat{\varphi}(x)}{\nu\sigma(x)} \geq \Psi^{-1}(1 - \epsilon).$$

Notice that, since $f_{\bar{\varphi}(x)/\nu}$ is symmetric around the origin, then $\Psi^{-1}(1 - \epsilon)$ is non-negative iff $\epsilon \leq 0.5$. Thus, defining

$$\kappa_{\epsilon,r} \doteq \nu \Psi^{-1}(1 - \epsilon), \quad \epsilon \in (0, 0.5] \quad (2.19)$$

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

$$\kappa_{\epsilon,r} \sigma(x) + \hat{\varphi}(x) \leq 0. \quad (2.20)$$

Recalling definitions (1.10), (1.11), we conclude that (2.20) is a convex constraint on x , and in particular, it is a second order cone (SOC) convex constraint, see e.g. [15]. We have therefore proved the following theorem.

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

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

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

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

where $\kappa_{\epsilon,r} = \nu\Psi^{-1}(1 - \epsilon)$, being Ψ the cumulative probability function of the density (2.18), and ν given by (2.16). $\triangleleft$

In some cases of interest, the cumulative distribution, and hence the corresponding safety parameter, can be computed in closed form. This is for instance the case for the Gaussian and the uniform distribution over ellipsoidal support, which are considered next.

2.1 Gaussian distribution

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

$$g(r) = \frac{1}{(2\pi)^{v/2}} \exp(-r^2/2).$$

Consequently, $f_{\hat{\varphi}(x)/\nu}$ from (2.18) is the Gaussian density function

$$f_{\hat{\varphi}(x)/\nu}(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-\xi^2/2),$$

and Ψ is the standard Gaussian cumulative probability function

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

For $\epsilon \in (0, 0.5]$, the safety parameter κ_ϵ is therefore given by $\kappa_\epsilon = \kappa_{\epsilon,G} = \Psi_G^{-1}(1 - \epsilon)$. Notice that we here recover a classical result (see [20], Theorem 10.4.1), that could also be derived more directly, without passing through the radial densities framework.

2.2 Uniform distribution on ellipsoidal support

The following lemma holds.

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 $Q \doteq \nu\Gamma_f$, $\Gamma \succ 0$, and $\nu \doteq \sqrt{n+3}$. Then, for any $\epsilon \in (0, 0.5]$, the chance constraint*

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

is equivalent to the convex second order cone constraint

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

where

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

being $\Psi_{\text{Beta}}(\cdot)$ the cumulative distribution of a $\text{Beta}(\frac{1}{2}; \frac{n}{2} + 1)$ probability density. $\triangleleft$

Proof. We first observe 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 (non-singular 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.

$$f_\omega(\omega) = g(\|\omega\|) = \begin{cases} \frac{1}{V_{n+1}} & \text{if } \|\omega\| \leq 1 \\ 0 & \text{otherwise,} \end{cases} \quad (2.21)$$

and V_{n+1} denotes the volume of the Euclidean ball of unit radius in $\mathbb{R}^{n+1}$.¹ It follows that d is Q -radial with $Q = \nu\Gamma_f$ and defining function $g(\cdot)$ given in (2.21). Notice that this specific choice of parameter ν is made according to (2.16), in order to fix the covariance of d to be equal to Γ . With this defining function, we can explicitly solve the integral in (2.18), obtaining the density

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

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

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

With a change of variable in the integral $z = t^2$ we finally obtain

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

where Ψ_{Beta} denotes the cumulative distribution of a Beta $(\frac{1}{2}; \frac{n}{2} + 1)$ probability density (see for instance [17] for a definition of Beta densities). The statement of the lemma then follows applying Theorem 2.1, where for $\epsilon \in (0, 0.5]$ the safety parameter κ_ϵ is given by

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

□

Remark 2.3. Contrary to the Gaussian case, the safety parameter for the uniform distribution on ellipsoidal support depends on the problem dimension n . In Figure 1, the value of $\kappa_{\epsilon,G}$ as a function of ϵ is compared to that of $\kappa_{\epsilon,u}$, for different values of n . Notice that, the covariance matrices being equal, the safety parameter $\kappa_{\epsilon,u}$ relative to the uniform distribution is essentially below (for small ϵ) 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 [4, 5] one seeks to satisfy the constraints for *all possible* values of some uncertain but bounded parameters, while 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.

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

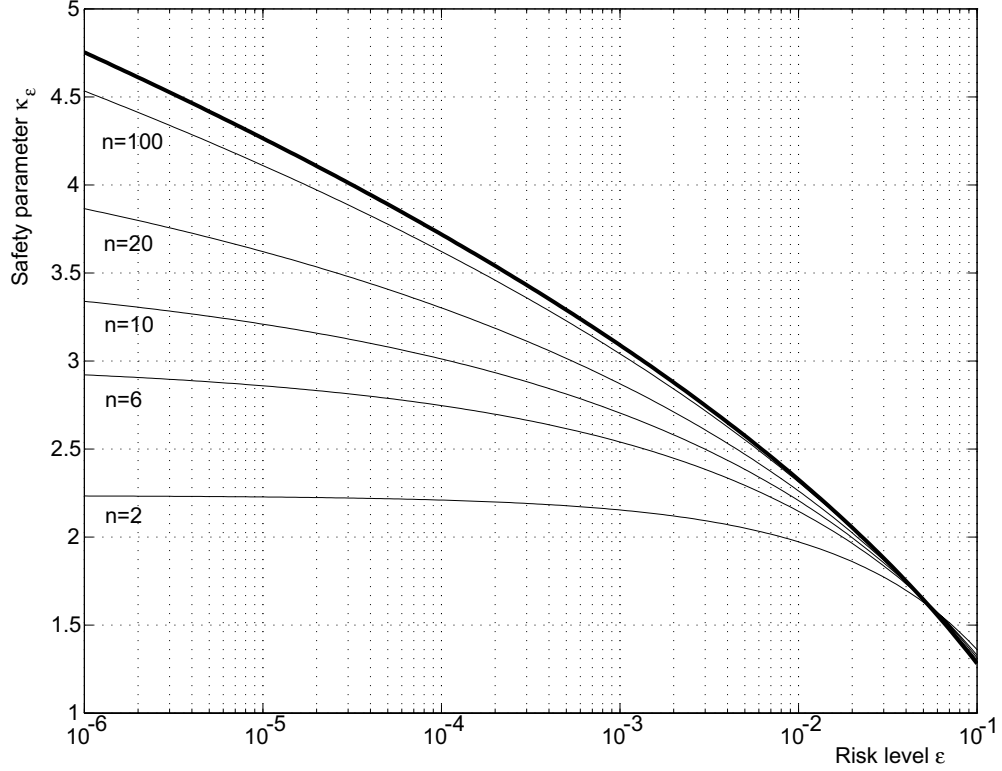


Figure 1: Comparison between $\kappa_{\epsilon,G}$ (thick line) and $\kappa_{\epsilon,u}$ (light lines), for $\epsilon \in [10^{-6}, 10^{-1}]$, and various values of n .

Consider a linear constraint $d^T \tilde{x} \leq 0$, where the data d lies in the ellipsoid $\mathcal{E} = \{\hat{d} + \kappa \Gamma_f z : \|z\| \leq 1\}$, $\Gamma \succ 0$, and suppose 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

$$d^T \tilde{x} \leq 0, \quad \forall d \in \mathcal{E} \quad (2.22)$$

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 of the form (2.22), for some suitable value of κ , and vice-versa. $\star$

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 non-increasing probability densities.

3 Distributionally Robust Chance Constraints

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

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

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

$$\inf_{d \sim \mathcal{D}} \text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon,$$

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, while 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 non-increasing distributions introduced in [1, 2].

3.1 Distributions with known mean and covariance

The first problem 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*

$$\inf_{d \sim (\hat{d}, \Gamma)} \text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \quad (3.23)$$

is equivalent to the convex second order cone constraint

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

with

$$\kappa_\epsilon = \sqrt{\frac{1 - \epsilon}{\epsilon}}. \quad (3.24)$$

◁

Proof. We express first d as $d = \hat{d} + \Gamma_f z$, where $E\{z\} = 0$, $\text{var}\{z\} = I$, and consider initially the case (a) when $\Gamma_f^T \tilde{x} \neq 0$. Then, from a classical result of Marshall and Olkin [16] (see also [6], Theorem 9), we have that

$$\sup_{d \sim (\hat{d}, \Gamma)} \text{Prob} \{d^T \tilde{x} > 0\} = \sup_{z \sim (0, I)} \text{Prob} \{z^T \Gamma_f^T \tilde{x} > -\hat{d}^T \tilde{x}\} = \frac{1}{1 + q^2},$$

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 notice that if $\tilde{x}^T \hat{d} > 0$, 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

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

Summarizing, we have

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

hence, the constraint (3.23) is satisfied if and only if

$$\frac{1}{1+q^2} \leq \epsilon,$$

i.e.

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

or

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

which proves that, in case (a), (3.23) is equivalent to (3.24).

On the other hand, in the case (b) when $\Gamma_f^T \tilde{x} = 0$, we simply have that $\inf_{d \sim (\hat{d}, \Gamma)} \text{Prob}\{d^T \tilde{x} \leq 0\} = 1$, if $\hat{\varphi}(x) \leq 0$, and it is zero otherwise. Therefore, since $\sigma(x) = 0$, it follows that (3.23) is still equivalent to (3.24), which concludes the proof. $\square$

A result analogous to Theorem 3.1 has been recently employed in [14] in the specific context of kernel methods for classification. We next discuss 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 $A \subseteq \mathbb{R}^n$. $\triangleleft$

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 satisfaction of the chance constraint, robustly over the family $(\hat{d}, \Gamma)_\mathcal{S}$.

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

$$\inf_{d \sim (\hat{d}, \Gamma)_\mathcal{S}} \text{Prob}\{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \tag{3.25}$$

holds if

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

with

$$\kappa_\epsilon = \sqrt{\frac{1}{2\epsilon}}. \tag{3.26}$$

$\triangleleft$

Proof. If d is symmetric around $\hat{d}$, then for all x , $\varphi(x)$ is symmetric around $\hat{\varphi}(x)$. Therefore (3.25) is satisfied if

$$\sup_{\varphi(x) \sim (\hat{\varphi}(x), \sigma^2(x))_S} \text{Prob} \{ \varphi(x) > 0 \} \leq \epsilon. \quad (3.27)$$

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

$$\sup_{z \sim (E\{z\}, \text{var}\{z\})_S} \text{Prob} \{ z > a \} = \begin{cases} \frac{1}{2} \min \left\{ 1, \frac{\text{var}\{z\}}{(E\{z\} - a)^2} \right\} & \text{if } a < E\{z\} \\ 1 & \text{otherwise.} \end{cases}$$

In our case, we hence have that for $\epsilon \in (0, 0.5]$, $\sup_{\varphi(x) \sim (\hat{\varphi}(x), \sigma^2(x))_S} \text{Prob} \{ \varphi(x) > 0 \} \leq \epsilon$ is satisfied whenever $\sigma^2(x) \leq 2\epsilon\hat{\varphi}(x)$ and $\hat{\varphi}(x) < 0$, i.e. when

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

from which the statement immediately follows. $\square$

3.2 Random data in independent intervals

In this section, we analyze a data uncertainty model where the random data d has known mean $\hat{d}$, and its individual elements are only known to belong with probability one to independent bounded intervals, i.e. we assume

$$d_i = \hat{d}_i + \omega_i, \quad i = 1, \dots, n+1 \quad (3.28)$$

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 $(n+1)$ -dimensional random variable d satisfying the above definition, where L is a diagonal matrix containing the interval widths

$$L \doteq \text{diag}(\ell_1^+ - \ell_1^-, \dots, \ell_{n+1}^+ - \ell_{n+1}^-). \quad (3.29)$$

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*

$$\inf_{d \sim (\hat{d}, L)_{\mathcal{I}}} \text{Prob} \{ d^T \tilde{x} \leq 0 \} \geq 1 - \epsilon \quad (3.30)$$

holds if

$$\sqrt{\frac{1}{2} \ln \frac{1}{\epsilon}} \|L\tilde{x}\| + \hat{\varphi}(x) \leq 0. \quad (3.31)$$

$\triangleleft$

Proof. By definition, any random vector d whose density belongs to the class $(\hat{d}, L)_{\mathcal{I}}$ is expressed by (3.28), and

$$d^T \tilde{x} = \hat{d}^T \tilde{x} + \sum_{i=1}^{n+1} \xi_i, \quad (3.32)$$

where we defined

$$\xi_i \doteq x_i \omega_i, \quad i = 1, \dots, n$$

and $\xi_{n+1} \doteq \omega_{n+1}$. With our usual notation $\varphi(x) = d^T \tilde{x}$ and $\hat{\varphi}(x) = \hat{d}^T \tilde{x}$, we hence have that

$$\text{Prob}\{\varphi(x) \leq 0\} = \text{Prob}\left\{\sum_{i=1}^{n+1} \xi_i \leq -\hat{\varphi}(x)\right\}. \quad (3.33)$$

Now, the ξ_i 's 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 Hoeffding's [11] tail probability inequality to (3.33), we obtain that, if $\hat{\varphi}(x) \leq 0$ then

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

from which the statement easily follows. $\square$

3.3 Radially symmetric non-increasing distributions

In this section, we consider two classes of radially symmetric non-increasing distributions (RSNID) whose supports are defined by means of the Euclidean and infinity norms. These are special cases of the RSNIDs introduced in [1, 2], 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

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

where $P = \text{diag}(p_1, \dots, p_{n+1}) \succ 0$, $Q \succ 0$. Clearly, $\mathcal{H}(\hat{d}, P)$ is an orthotope centered in $\hat{d}$ and 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

$$d - E\{d\} = P\omega \quad \text{resp.} \quad d - E\{d\} = Q\omega$$

where ω is a random vector having probability density f_ω such that

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

and where $g(\cdot)$ is a non-increasing function. $\triangleleft$

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 [1], 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*

$$\inf_{d \sim \mathcal{F}_{\mathcal{H}}} \text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \quad (3.34)$$

is equivalent to the chance constraint

$$\text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad d \sim U(\mathcal{H}(\hat{d}, P)) \quad (3.35)$$

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

$$\inf_{d \sim \mathcal{F}_{\mathcal{E}}} \text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \quad (3.36)$$

is equivalent to the chance constraint

$$\text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon, \quad d \sim U(\mathcal{E}(\hat{d}, Q)) \quad (3.37)$$

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

The proof of this proposition is readily established from Theorem 6.3 of [2]. Indeed, it suffices to show (we consider here only the family $\mathcal{F}_{\mathcal{E}}$, the case of $\mathcal{F}_{\mathcal{H}}$ 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 $\mathcal{F}_{\mathcal{E}}$, $\text{Prob} \{d^T \tilde{x} \leq 0\} \geq 0.5$ 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 $\tilde{x}^T Q \omega \leq -\hat{d}^T \tilde{x}$ implies $\rho \tilde{x}^T Q \omega \leq \rho(-\hat{d}^T \tilde{x})$, and 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 (3.37), 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 (3.35) 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 (3.35) to hold. In this respect, we notice that it was already shown in [12] that for $\epsilon \in (0, 0.5]$, $\text{Prob}_{d \sim U(\mathcal{H}(\hat{d}, P))} \{d^T \tilde{x} \leq 0\}$ is a convex constraint on x . However, no explicit condition was given in [12] to enforce such convex constraint. At 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 explicitly converted 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. $\star$

Lemma 3.3. *For any $\epsilon \in (0, 0.5]$, the distributionally robust chance constraint (3.34) holds if*

$$\sqrt{\frac{1}{6} \ln \frac{1}{\epsilon}} \|2P\tilde{x}\| + \hat{\varphi}(x) \leq 0. \quad (3.38)$$

$\triangleleft$

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

$$\ln E\{e^{\lambda\xi}\} \leq \frac{\lambda^2 c^2}{6}. \quad (3.39)$$

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

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

and consider the function $\psi(z) \doteq \ln \frac{\sinh(z)}{z}$, $z \doteq \lambda c$, 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 the first and second derivatives of ψ , respectively, and moreover $\psi''(z) \leq 1/3$, $\forall z$. Therefore, by Taylor expansion with Lagrange remainder we have that for some $\theta \in [0, z]$

$$\psi(z) = \psi(0) + z\psi'(0) + \frac{z^2}{2}\psi''(\theta) \leq \frac{z^2}{6},$$

from which (3.39) immediately follows.

Now, we write $d = \hat{d} + P\omega$ (we recall that $P = \text{diag}(p_1, \dots, p_{n+1}) \succ 0$), and follow the same steps of the proof of Lemma 3.2 up to (3.33). 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 's appearing in (3.33) 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 $\hat{\varphi}(x) \leq 0$ and any $\lambda \geq 0$

$$\text{Prob} \left\{ \sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x) \right\} \leq \frac{E \left\{ e^{\lambda \sum_{i=1}^{n+1} \xi_i} \right\}}{e^{-\lambda \hat{\varphi}(x)}} = \frac{\prod_{i=1}^{n+1} E \left\{ e^{\lambda \xi_i} \right\}}{e^{-\lambda \hat{\varphi}(x)}}.$$

By (3.39) we further have

$$E \left\{ e^{\lambda \xi_i} \right\} \leq e^{\frac{(\lambda p_i x_i)^2}{6}}, \quad i = 1, \dots, n, \quad \text{and} \quad E \left\{ e^{\lambda \xi_{n+1}} \right\} \leq e^{\frac{(\lambda p_{n+1})^2}{6}}$$

and hence

$$\text{Prob} \left\{ \sum_{i=1}^{n+1} \xi_i > -\hat{\varphi}(x) \right\} \leq e^{\lambda^2 \|2P\tilde{x}\|^2 / 24 + \lambda \hat{\varphi}(x)} \leq e^{-6 \frac{\hat{\varphi}^2(x)}{\|2P\tilde{x}\|^2}} \quad (3.40)$$

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

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

Finally, the probability on the LHS of (3.40) is smaller than $\epsilon \in (0, 0.5]$ if $\hat{\varphi}(x) \leq 0$ and

$$\|2P\tilde{x}\|^2 \ln \frac{1}{\epsilon} \leq 6\hat{\varphi}^2(x),$$

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

$$\sqrt{\frac{1}{6} \ln \frac{1}{\epsilon}} \|2P\tilde{x}\| + \hat{\varphi}(x) \leq 0,$$

thus proving the claim. $\square$

Remark 3.2. Notice that the result in Lemma 3.3 improves (decreases) by a $\sqrt{3}$ factor 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 non-symmetric and possibly increasing densities, and it is hence richer than the class $\mathcal{F}_{\mathcal{H}}$ considered in Lemma 3.3. $\star$

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

$$\begin{aligned}\hat{d}_N &= \frac{1}{N} \sum_{i=1}^N d^i \\ \Gamma_N &= \frac{1}{N} \sum_{i=1}^N (d^i - \hat{d}_N)(d^i - \hat{d}_N)^T.\end{aligned}$$

Clearly, care should be now exerted, 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 [21] 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 (Theorem 3 of [21]). *Let $d^1, \dots, d^N \in \mathbb{R}^{n+1}$ be an N -sample generated independently at random according to an (unknown) distribution Π , and let $R = \sup_{d \in \text{support}(\Pi)} \|d\|$, $\beta \in (0, 1)$. Then with probability at least $1 - \beta$ it holds that*

$$\|\hat{d}_N - \hat{d}\| \leq \frac{R}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{1}{\beta}} \right). \quad (4.41)$$

$\triangleleft$

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

Lemma 4.2 (Corollary 2 of [21]). *Let $d^1, \dots, d^N \in \mathbb{R}^{n+1}$ be an N -sample generated independently at random according to an (unknown) distribution Π , and let $R = \sup_{d \in \text{support}(\Pi)} \|d\|$, $\beta \in (0, 1)$. Then, provided that*

$$N \geq \left(2 + \sqrt{2 \ln \frac{2}{\beta}} \right)^2,$$

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

$$\|\Gamma_N - \Gamma\|_F \leq \frac{2R^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{2}{\beta}} \right). \quad (4.42)$$

$\triangleleft$

We next present 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 $d^1, \dots, d^N$ be N independent samples of the random vector $d \in \mathbb{R}^{n+1}$ having an (unknown) distribution Π , 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, and denote with $\hat{d}, \Gamma$ the respective true (unknown) values. Define also*

$$\begin{aligned} r_1 &\doteq \frac{R}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{2}{\delta}} \right) \\ r_2 &\doteq \frac{2R^2}{\sqrt{N}} \left(2 + \sqrt{2 \ln \frac{4}{\delta}} \right). \end{aligned}$$

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

$$\inf_{d \sim (\hat{d}, \Gamma)} \text{Prob} \{d^T \tilde{x} \leq 0\} \geq 1 - \epsilon \quad (4.43)$$

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

$$N \geq \left(2 + \sqrt{2 \ln \frac{4}{\delta}} \right)^2$$

and

$$\sqrt{\frac{1-\epsilon}{\epsilon}} \sqrt{\tilde{x}^T (\Gamma_N + r_2 I) \tilde{x}} + \hat{d}_N^T \tilde{x} + \|\tilde{x}\| r_1 \leq 0. \quad (4.44)$$

◁

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

$$\hat{d} = \hat{d}_N + \xi, \quad \text{for some } \xi \in \mathbb{R}^{n+1} : \|\xi\| \leq r_1. \quad (4.45)$$

Similarly, applying Lemma 4.2 with $\beta = \delta/2$, we have that,

$$\Gamma = \Gamma_N + \Delta, \text{ for some } \Delta \in \mathbb{R}^{n+1, n+1} : \|\Delta\|_F \leq r_2, \quad (4.46)$$

holds with probability at least $1 - \delta/2$, provided that

$$N \geq \left(2 + \sqrt{2 \ln \frac{4}{\delta}} \right)^2.$$

Combining the above two events, we have that equations (4.45), (4.46) jointly hold with probability at least $(1 - \delta/2)^2 \geq 1 - \delta$.

Now, from Theorem 3.1 we know that

$$\kappa_\epsilon \sqrt{\tilde{x}^T \Gamma \tilde{x}} + \tilde{d}^T \tilde{x} \leq 0, \quad \kappa_\epsilon = \sqrt{\frac{1-\epsilon}{\epsilon}} \quad (4.47)$$

implies the satisfaction of the chance constraint (4.43). Hence, substituting (4.45), (4.46) in (4.47) we have that (4.43) holds with probability at least $1 - \delta$, provided that

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

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

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

□

5 Applications

5.1 Portfolio optimization

Consider a portfolio $p^- \in \mathbb{R}^n$ consisting of n assets (which may include cash), where $p_i^- \geq 0$, $i = 1, \dots, n$ represent the dollar value of the current holdings in each asset. In the so-called single-stage portfolio problem, the investor should decide an adjustment $x \in \mathbb{R}^n$ of the portfolio, where x_i denotes the dollar amount transacted in asset i ($x_i > 0$ for buying, and $x_i \leq 0$ for selling), and after the transaction the adjusted portfolio $p = p^- + x$ is held for a fixed amount of time.

In the classical Markowitz framework, the investor goal is to maximize the total expected wealth at the end of the investment period, while satisfying a set of constraints on the portfolio, which may include bounds on the amounts held in each single asset, as well as limits on the exposure to risk.

A standard approach to the problem is to assume that the *return* r_i of asset i over the considered period is a random variable, and that the expected value $\hat{r}_i$ and covariance terms of the returns are known. We denote with $r \doteq [r_1 \ \dots \ r_n]^T$ the vector of returns, with $\hat{r}$ its expected value, and with Σ its covariance, and consider the family $(\hat{r}, \Sigma)$ of all possible distributions on the returns, compatible with the given mean and covariance.

Now, defining $R \doteq \text{diag}(r_1, \dots, r_n)$, we have that the portfolio at the end of the investment period is the random vector

$$p^+ = R(p^- + x),$$

while the total wealth is $\mathbf{1}^T p^+$, where $\mathbf{1}$ denotes a vector of ones.

We here consider the following simplified portfolio optimization problem:

$$\max_x E\{\mathbf{1}^T p^+\} \text{ subject to} \quad (5.48)$$

$$\mathbf{1}^T x + \tau(x) \leq 0 \quad (5.49)$$

$$p^- + x \geq 0 \quad (5.50)$$

$$\sup_{r \sim (\hat{r}, \Sigma)} \text{Prob}\{\mathbf{1}^T p^+ \leq w_{low}\} \leq \epsilon \quad (5.51)$$

where (5.49) is a budget constraint taking into account all transaction costs $\tau(x)$, (5.50) is a no-shortselling constraint, and (5.51) is a shortfall risk constraint. In words, this latter constraint imposes a small probability ϵ on the event that the end-of-period wealth be lower than an undesired level w_{low} . Notice that we do not assume a specific distribution for the returns, but rather impose the probability constraint robustly with respect to all probability distributions on the returns that have the specified mean and covariances. In our schematic example, the transaction costs are assumed to be proportional to the transacted amounts, i.e.

$$\tau(x) \doteq \alpha \sum_{i=1}^n |x_i| = \alpha \|x\|_1,$$

where $\alpha \geq 0$ is the fixed unit transaction cost.

Now, the probability constraint (5.51) can be converted into an explicit convex constraint applying Theorem 3.1. Specifically, the distributionally robust constraint (5.51) holds if

$$\kappa \|\Sigma_f(p^- + x)\| - \hat{r}^T(p^- + x) + w_{low} \leq 0,$$

with $\kappa = \sqrt{(1 - \epsilon)/\epsilon}$, and Σ_f full-rank, such that $\Sigma_f \Sigma_f^T = \Sigma$. Therefore, the stochastic problem (5.48)–(5.51) is converted into the explicit convex problem

$$\begin{aligned} \max_x \quad & \hat{r}^T(p^- + x) \text{ subject to} \\ & \mathbf{1}^T x + \alpha \|x\|_1 \leq 0 \\ & p^- + x \geq 0 \\ & \kappa \|\Sigma_f(p^- + x)\| - \hat{r}^T(p^- + x) + w_{low} \leq 0. \end{aligned} \quad (5.52)$$

For the purpose of the example, we considered a portfolio holding period of 20 days, and five assets from the S&P 500 basket (tickers: AOL, CSCO, DELL, EQR, TXN), plus cash, i.e. $n = 6$. Cash is assumed to have unit return and zero covariance (riskless asset). We (crudely) estimated the 20-day average returns and covariances for the assets from historical data (closing prices from 2002-05-14 to 2003-05-13, using 0.98 forgetting factor), obtaining

$$\hat{r}^T = \begin{bmatrix} 1.2018 & 1.2197 & 1.1744 & 1.0698 & 1.3776 & 1 \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} 0.5014 & 0.1839 & 0.1471 & 0.0616 & 0.2354 & 0 \\ 0.1839 & 0.4817 & 0.2350 & 0.0891 & 0.4921 & 0 \\ 0.1471 & 0.2350 & 0.4868 & 0.0594 & 0.3275 & 0 \\ 0.0616 & 0.0891 & 0.0594 & 0.0733 & 0.1587 & 0 \\ 0.2354 & 0.4921 & 0.3275 & 0.1587 & 1.8530 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Assuming transaction cost $\alpha = 1\%$, initial holdings $p^- = [1\ 1\ 1\ 1\ 1\ 1]^T$ and $w_{low} = 4$, we obtained the updated portfolios shown in Figure 2, for different values of risk level ϵ . Notice that the use of a distributionally robust probability constraint leads to very ‘cautious’ choices of the portfolios (strong bias towards riskless asset). More ‘aggressive’ portfolios are instead obtained if we assume that the random returns obey to a Gaussian distribution, in which case the constant κ in (5.52) is set to $\kappa = \Psi_G^{-1}(1 - \epsilon)$, where Ψ_G is the standard Gaussian cumulative function.

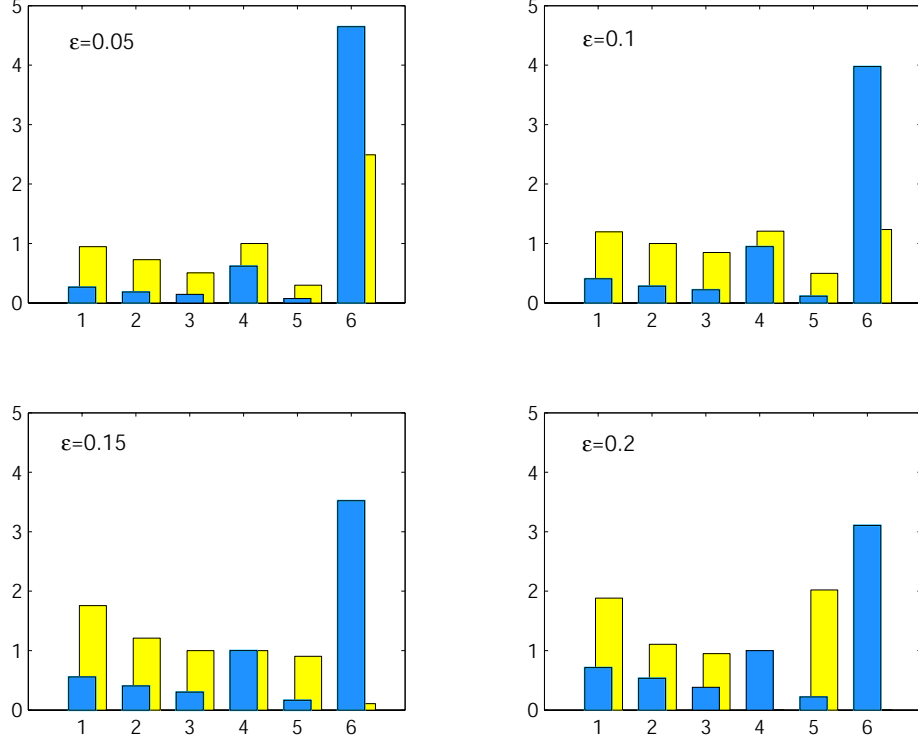


Figure 2: Composition of optimal portfolios for different values of risk level ϵ . The abscissae report the asset type (1=AOL, 2=CSCO, 3=DELL, 4=EQR, 5=TXN, 6=cash). The darker bars show the composition of the distributionally robust portfolios, while the light ones refer to the Gaussian case.

5.2 Model predictive control with process disturbance

Consider a linear time invariant model of the form

$$x(k+1) = Ax(k) + B_u u(k) + B_v v(k) \quad (5.53)$$

$$y(k) = Cx(k) \quad (5.54)$$

where $x \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^{n_u}$ is the control input, $y \in \mathbb{R}^{n_y}$ is the system output, and $v \in \mathbb{R}^{n_v}$ is an exogenous random disturbance. In model predictive control (also known as receding horizon control), a stabilizing control law u is sought for this system, such that some assigned constraints on the control effort and on the outputs are fulfilled. In particular, model

predictive control (MPC) achieves these goals by solving iteratively over a sliding horizon window an optimization problem of the type

$$\min_{U_k^{k+N-1}} J(U_k^{k+N-1}, x(k)) \quad \text{subject to:} \quad (5.55)$$

$$y_{min} \leq y(k+i|k) \leq y_{max}, \quad i = 1, \dots, N \quad (5.56)$$

$$u_{min} \leq u(k+i) \leq u_{max}, \quad i = 0, \dots, N-1 \quad (5.57)$$

$$x(k+i+1|k) = Ax(k+i|k) + B_u u(k+i), \quad i \geq 0, \quad x(k|k) = x(k) \quad (5.58)$$

$$y(k+i) = Cx(k+i|k), \quad i \geq 0, \quad (5.59)$$

where $x(k+i|k)$ denotes the prediction of the state at time $k+i$ done according to the model (5.53) on the basis of the knowledge of the system state $x(k)$, and *neglecting the disturbance* $v(k)$. The decision variable of the above problem is the sequence of input actions $U_k^{k+N-1} = \{u(k), u(k+1), \dots, u(k+N-1)\}$, and the minimization objective is usually a quadratic cost function of the form

$$J(U_k^{k+N-1}, x(k)) = x^T(k+N|k)Px(k+N|k) + \sum_{i=0}^{N-1} (x^T(k+i|k)Qx(k+i|k) + u^T(k+i)Ru(k+i)),$$

where P, Q, R are assigned positive definite matrices. The MPC control law prescribes to actually apply to the system only the first component $u^*(k)$ of the optimal input sequence resulting from problem (5.55), and then to run the optimization again at time $k+1$, using the available measurement of the new state $x(k+1)$, and so on in a sliding window fashion.

Notice however that the prediction model in constraint (5.58) does not take into consideration that fact that the actual plant model (5.53) is affected by a random disturbance $v(k)$. This has two consequences: First, there is in reality no guarantee that the system output would satisfy the desired bounds (5.56), and second the objective cost does not take into account the randomness present in the states.

To overcome these difficulties, we here propose to consider the true noisy model in (5.58). This implies in turn that the constraints (5.56) should be imposed *up to a given level of probability*, and that an expectation is to be introduced in the objective J . While considering next this probabilistic setup, we wish to make it clear that it is not within the scope of this presentation to address issues of asymptotic stability of the resulting policy, when the control law is applied in a sliding horizon fashion. Hence, we here concentrate on the solution the finite-horizon open-loop optimization problem (5.55).

Given the state $x(k|k) = x(k)$, the state at time $k+i$ predicted according to model (5.53)–(5.54) is the random vector

$$x(k+i|k) = A^i x(k) + \Phi_{u,i} U_k^{k+i-1} + \Phi_{v,i} V_k^{k+i-1},$$

and the j -th component of the corresponding predicted output is given by

$$y_j(k+i|k) = C_j A^i x(k) + C_j \Phi_{u,i} U_k^{k+i-1} + C_j \Phi_{v,i} V_k^{k+i-1}, \quad (5.60)$$

where C_j denotes the j -th row of matrix C , and where we defined

$$\begin{aligned}\Phi_{u,i} &\doteq \begin{bmatrix} A^{i-1}B_u & \cdots & AB_u & B_u \end{bmatrix}, \quad U_k^{k+i-1} \doteq \begin{bmatrix} u(k) \\ \vdots \\ u(k+i-1) \end{bmatrix}; \\ \Phi_{v,i} &\doteq \begin{bmatrix} A^{i-1}B_v & \cdots & AB_v & B_v \end{bmatrix}, \quad V_k^{k+i-1} \doteq \begin{bmatrix} v(k) \\ \vdots \\ v(k+i-1) \end{bmatrix}.\end{aligned}$$

Now, let $v(k)$ be zero-mean independent random vectors, with covariance $\Sigma(k)$. Then, the optimization objective is explicitly written as

$$\begin{aligned}J_E(U_k^{k+N-1}, x(k)) &\doteq E\{J(U_k^{k+N-1}, x(k))\} = \\ &(A^N x(k) + \Phi_{u,N} U_k^{k+N-1})^T P (A^N x(k) + \Phi_{u,N} U_k^{k+N-1}) + \\ &\text{trace}(\Phi_{v,N}^T P \Phi_{v,N} \text{diag}(\Sigma(k), \dots, \Sigma(k+N-1))) + \\ &\sum_{i=0}^{N-1} \left((A^i x(k) + \Phi_{u,i} U_k^{k+i-1})^T Q (A^i x(k) + \Phi_{u,i} U_k^{k+i-1}) + \right. \\ &\left. \text{trace}(\Phi_{v,i}^T Q \Phi_{v,i} \text{diag}(\Sigma(k), \dots, \Sigma(k+i-1))) + u^T(k+i) R u(k+i) \right).\end{aligned}$$

Notice incidentally that $J_E(U_k^{k+N-1}, x(k))$ differs from $J(U_k^{k+N-1}, x(k))$ only for an additive term that is independent of the decision variable U_k^{k+N-1} , and therefore the presence of noise has in fact no effect on the function to be minimized.

The probabilistic version of problem (5.55) is

$$\min_{U_k^{k+N-1}} J_E(U_k^{k+N-1}, x(k)) \text{ subject to:} \quad (5.61)$$

$$\text{Prob}\{y_j(k+i|k) \leq y_{\max,j}\} \geq 1 - \epsilon, \quad i = 1, \dots, N; j = 1, \dots, n_y \quad (5.62)$$

$$\text{Prob}\{y_j(k+i|k) \geq y_{\min,j}\} \geq 1 - \epsilon, \quad i = 1, \dots, N; j = 1, \dots, n_y \quad (5.63)$$

$$u_{\min} \leq u(k+i) \leq u_{\max}, \quad i = 0, \dots, N-1, \quad (5.64)$$

where $y_j(k+i|k)$ is given by (5.60). For the purpose of this example, we consider the disturbances $v(k)$ s to be zero-mean independent and uniform in the interval $[-p, p]$. Notice that, by the uniformity principle, this also corresponds to considering distributional robust probability constraints over distributions belonging to the symmetric non-increasing class discussed in Section 3.3.

Next, notice that each of the (5.62) constraints (or the (5.63) constraints) has the generic form

$$\text{Prob}\{a^T U + b \leq 0\} \geq 1 - \epsilon, \quad \text{with } b \doteq \beta^T V + \gamma \quad (5.65)$$

where $U \in \mathbb{R}^h$ is the decision vector, $V \in \mathbb{R}^h$ is the noise vector composed of independent bounded elements, and $a, \beta \in \mathbb{R}^h$, $\gamma \in \mathbb{R}$ are fixed vectors. We have therefore a simplified situation in which the vector a that multiplies the decision variable is not random. Indeed, in the present case only the additive term b is random, and it is convenient to exploit the fact that this term is a sum of

bounded random variables, which induces an averaging, or ‘concentration’ effect. Specifically, using (*mutatis mutandis*) inequality (3.40), we have that, for $a^T U + \gamma \leq 0$,

$$\text{Prob}\{a^T U + b > 0\} = \text{Prob}\{\beta^T V > -(a^T U + \gamma)\} \leq e^{-\frac{3}{2} \frac{(a^T U + \gamma)^2}{p^2 \|\beta\|^2}}.$$

Therefore, the probability constraint (5.65) holds for $\epsilon \in (0, 0.5]$, if

$$a^T U + \gamma + \sqrt{\frac{2}{3} \ln \frac{1}{\epsilon} p \|\beta\|} \leq 0. \quad (5.66)$$

In conclusion, in the considered setup and due to the additive nature of the disturbance, the probability constraints are simply translated into linear constraints. This in turns means that probabilistic guarantees against disturbances in MPC are obtained at no ‘extra cost,’ i.e. there is no increase of the computational complexity of the problem.

As a numerical test, we considered a modification of an example proposed in [3]. The system model is

$$\begin{aligned} x(k+1) &= \begin{bmatrix} 0.7326 & -0.0861 \\ 0.1722 & 0.9909 \end{bmatrix} x(k) + \begin{bmatrix} 0.0609 \\ 0.0064 \end{bmatrix} u(k) + \begin{bmatrix} 1 \\ 0.2 \end{bmatrix} v(k) \\ y(k) &= x(k). \end{aligned}$$

We consider horizon $N = 2$, weights $Q = I$, $R = 0.01$, and P that solves the Lyapunov equation $P = A^T P A + Q$. We assumed noise bound $p = 0.1$, initial state $x(0) = [1 \ 1]^T$, and the design constraints on input $-2 \leq u(k) \leq 2$, and on output $y(k) \geq [-0.3 \ -0.3]^T$. In this setting, Figure 3 shows the resulting state trajectories under receding horizon control law computed through problem (5.55), i.e. neglecting the presence of the disturbance.

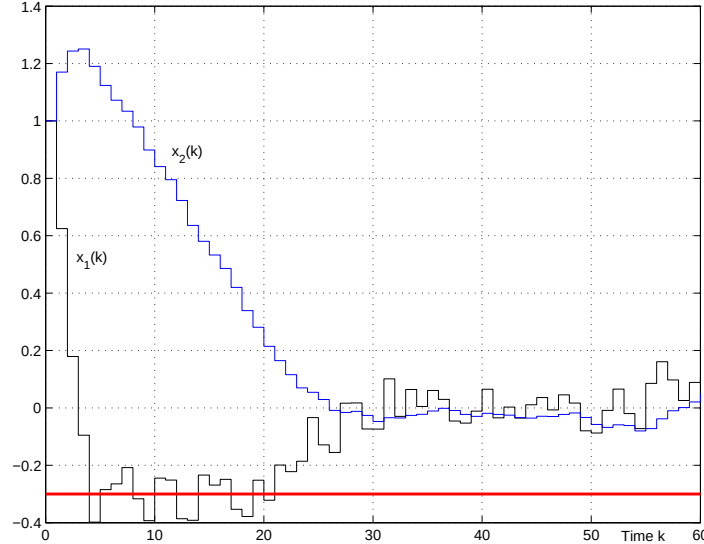


Figure 3: State trajectories under receding horizon control law obtained by solution of problem (5.55). The thick horizontal line marks the bound $y(k) \geq [-0.3 \ -0.3]^T$.

It is apparent from this figure that the actual state trajectories fail to satisfy the required bounds, at certain time instants. We then applied to the system the control law computed according to the probability constrained linear program, with constraints of type (5.66), setting risk level $\epsilon = 0.2$. The resulting state trajectories are shown in Figure 3, and, in this simulation, guarantee the satisfaction of the output constraints, despite the presence of the disturbance.

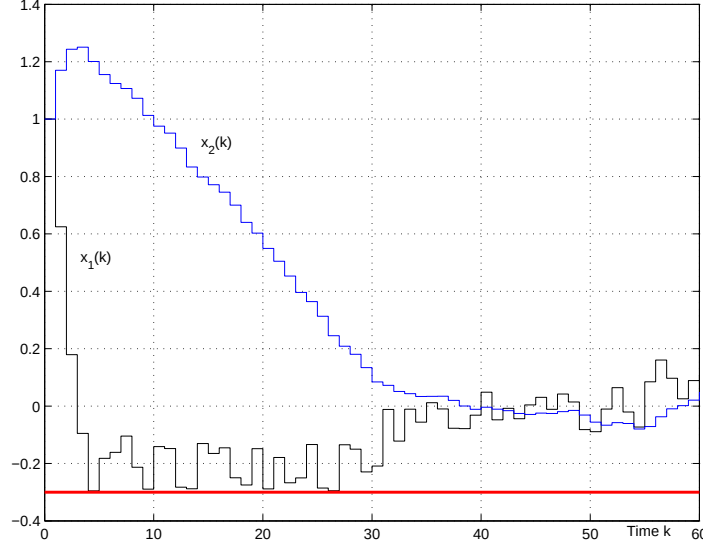


Figure 4: State trajectories under receding horizon control law obtained by solution of the chance-constrained optimization problem. The thick horizontal line marks the bound $y(k) \geq [-0.3 \ -0.3]^T$.

6 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 further analyzed 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 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 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. Finally, we illustrated some of the results with examples from optimal portfolio selection and model predictive control problems. Other examples of the use of chance constraints in control design have also recently appeared in [13].

References

- [1] B.R. Barmish and C.M. Lagoa. The uniform distribution: A rigorous justification for its use in robustness analysis. *Mathematics of Control, Signals, and Systems*, 10:203–222, 1997.
- [2] B.R. Barmish, C.M. Lagoa, and R. Tempo. Radially truncated uniform distributions for probabilistic robustness of control systems. In *Proc. American Control Conf.*, pages 853–857, Albuquerque, NM, June 1997.
- [3] A. Bemporad, M. Morari, V. Dua, and E.N. Pistikopoulos. The explicit quadratic regulator for constrained systems. *Automatica*, 38:3–20, 2002.
- [4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [5] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.
- [6] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. Technical report, Dept. Math. O.R., Cambridge, Mass 02139, 1998.
- [7] G. Calafiore and F. Dabbene. Loop gain under random feedback. In *Proc. IEEE Conf. on Decision and Control*, Orlando, FL, December 2001.
- [8] A. Charnes and W.W. Cooper. Deterministic equivalents for optimizing and satisfying under chance constraints. *Operations Research*, 11:18–39, 1963.
- [9] S.P. Uryasev (Ed.). *Probabilistic Constrained Optimization*. Kluwer Academic, The Netherlands, 2000.
- [10] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Monographs and surveys in pure and applied mathematics. Chapman & Hall - CRC, Boca Raton, 1999.
- [11] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.
- [12] C.M. Lagoa. On the convexity of probabilistically constrained linear programs. In *Proc. IEEE Conf. on Decision and Control*, pages 516–521, Phoenix, AZ, December 1999.
- [13] C.M. Lagoa, X. Li, and M. Sznaier. Application of probabilistically constrained linear programs to risk-adjusted controller design. In *Proc. American Control Conf.*, pages 738–743, Arlington, VA, June 2001.
- [14] G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M.I. Jordan. Minimax probability machine. In *Advances in Neural Information Processing Systems*, volume 14, 2001.
- [15] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, November 1998.

- [16] A. Marshall and I. Olkin. Multivariate Chebyshev inequalities. *Ann. Math. Stat.*, 31:1001–1014, 1960.
- [17] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
- [18] B.T. Polyak and P.S. Shcherbakov. Random spherical uncertainty in estimation and robustness. *IEEE Trans. Aut. Control*, 45(11):2145–2150, November 2000.
- [19] I. Popescu. An SDP approach to moment bounds for distributions with convex properties. Technical report, INSED working paper. Available on-line, 2002.
- [20] A. Prékopa. *Stochastic Programming*. Kluwer, 1995.
- [21] J. Shawe-Taylor and N. Cristianini. Estimating the moments of a random vector with applications. *Working Paper*, 2003.
- [22] S. Vajda. *Probabilistic Programming*. Academic Press, New York, 1972.
- [23] C. van de Panne and W. Popp. Minimum cost cattle feed under probabilistic problem constraint. *Management Science*, 9:405–430, 1963.