Randomized Strategies for Probabilistic Solutions of Uncertain Feasibility and Optimization Problems

Teodoro Alamo, *Member, IEEE*, Roberto Tempo, *Fellow, IEEE*, and Eduardo F. Camacho, *Senior Member, IEEE*

***Abstract*—In this paper, we study two general semi-infinite programming problems by means of a randomized strategy based on statistical learning theory. The sample size results obtained with this approach are generally considered to be very conservative by the control community. The first main contribution of this paper is to demonstrate that this is not necessarily the case. Utilizing as a starting point one-sided results from statistical learning theory, we obtain bounds on the number of required samples that are manageable for "reasonable" values of probabilistic confidence and accuracy. In particular, we show that the number of required samples grows with the accuracy parameter ϵ as $1/\epsilon \ln 1/\epsilon$, and this is a significant improvement when compared to the existing bounds which depend on $1/\epsilon^2 \ln 1/\epsilon^2$. Secondly, we present new results for optimization and feasibility problems involving Boolean expressions consisting of polynomials. In this case, when the accuracy parameter is sufficiently small, an explicit bound that only depends on the number of decision variables, and on the confidence and accuracy parameters is presented. For convex optimization problems, we also prove that the required sample size is inversely proportional to the accuracy for fixed confidence. Thirdly, we propose a randomized algorithm that provides a probabilistic solution circumventing the potential conservatism of the bounds previously derived.**

***Index Terms*—Probabilistic robustness, randomized algorithms, statistical learning theory, uncertain systems.**

I. Introduction

The presence of uncertainty in the system description has been always recognized as a critical issue in control theory and applications. Since the early 1980s, several approaches based on a direct characterization of the uncertainty into the plant have been proposed. The design objective hence becomes the computation of a controller that is guaranteed to perform satisfactorily against all possible uncertainty realizations, thus leading to a worst-case (or robust) solution. Typically, for a robustness problem, the design parameters, along with different auxiliary variables, are parameterized through the use of a decision variable vector θ, which is denoted as "design parameter," and is restricted to a design set Θ. On the other hand, the uncertainty is characterized by means of a given bounded set $\mathcal{W}$. That is, each element $w \in \mathcal{W}$ represents one of the admissible uncertainty realizations, and is called "uncertainty vector." However, worst-case control design problems are often intractable from a computational complexity point of view [10]. For this reason, various relaxation methods have been recently proposed, see for example the review paper [29]. Moreover, it is clear that such an approach would often lead to a rather conservative design. To avoid these issues, an alternative paradigm is to assume that the plant uncertainty is probabilistically described so that a randomized algorithm may be used to obtain, normally in polynomial time, a solution that satisfies some given properties [32], [37].

Uncertainty randomization is now widely accepted as an effective tool in dealing with control problems which are computationally difficult, see, e.g., [32]. In particular, regarding *synthesis* of a controller to achieve a given performance, two complementary approaches, sequential and non-sequential, have been proposed in recent years.

For sequential methods, the resulting iterative algorithms are based on stochastic gradient [14], [17], [20], [27], ellipsoid iterations [21], [25] or analytic center cutting plane methods [13], see also [5] for other classes of sequential algorithms. Convergence properties in finite-time are in fact one of the focal points of these papers. Various control problems have been solved using these sequential randomized algorithms, including robust LQ regulators, switched systems, and uncertain linear matrix inequalities (LMIs). Sequential methods are mostly used for convex problems; they are very useful because at each iteration the computational time is usually affordable. However, the number of iterations may be very large and depends on a stopping rule.

A classical approach for non-sequential methods is based upon statistical learning theory [24], [30], [34], and [37]. In particular, the use of this theory for feedback design of uncertain systems has been initiated in [38]; subsequent work along this direction include [39], [40]. However, the sample size bounds derived in these papers, which guarantee that the obtained solution meets a given probabilistic specification, may be too conservative for being practically useful in a systems and control context if the available computational resources are limited. The advantage of these methods is that the problem under attention may be non-convex and can be solved in one-shot without any stopping rule.

In particular, for convex optimization problems, a successful paradigm, denoted as the scenario approach, has been introduced in [11], [12]. In this approach, the original robust control

Manuscript received March 14, 2008; revised December 02, 2008. First published October 13, 2009; current version published November 04, 2009. This work was supported by the MCYT-Spain and the European Commission which funded this work under Projects DPI2007-66718-C04-01 and FP7-223866, respectively. Recommended by Associate Editor I. Paschalidis.

T. Alamo and E. F. Camacho are with the Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla, Sevilla 41092, Spain. (e-mail: alamo@cartuja.us.es; eduardo@cartuja.us.es).

R. Tempo is with IEIIT-CNR, Politecnico di Torino, Torino 10129, Italy. (e-mail: roberto.tempo@polito.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2009.2031207

problem is reformulated in terms of a single convex optimization problem with sampled constraints which are randomly generated. The main result of this line of research is to derive explicit bounds without resorting to statistical learning methods.

In this paper, we address two general *semi-infinite* problems (not necessarily convex) which encompass design of systems affected by uncertainty. These problems are called semi-infinite because they are subject to an infinite number of constraints, but the number of decision variables is finite. In particular, we consider a binary measurable function $g : \Theta \times \mathcal{W} \rightarrow \{0,1\}$ which serves to formulate the specific design problem under attention.

Semi-infinite feasibility problem: Find θ, if it exists, in the feasible set

$$\{\theta \in \Theta : g(\theta, w) = 0, \quad \text{for all } w \in \mathcal{W}\}. \tag{1}$$

Semi-infinite optimization problem: If the feasible set is nonempty, find the optimal solution of the problem

$$\min_{\theta\in\Theta} J(\theta) \text{ subject to } g(\theta, w) = 0, \quad \text{for all } w \in \mathcal{W} \tag{2}$$

where $J : \Theta \rightarrow (-\infty, \infty)$ is a measurable function.

In the recent randomization literature a distinction between optimization and feasibility is normally found. For example, papers [11] and [12] deal with optimization while [17] and [27] deal with feasibility. Also, in this paper there are results that apply only to optimization problems (see the convex scenario of Section VI) while others (more related to Statistical Learning Theory) do not rely on the assumption that an optimum has been reached.

We notice that robust stabilization, robust model predictive control and control of uncertain nonlinear systems, to number only a few, may be reformulated as (1) or (2). We also recall that various specific applications which require the solution of these general semi-infinite optimization problems include congestion and cooperative control, robust localization and trajectory planning. To provide further motivations we present two examples that fit within the proposed framework.

A. Example 1: Design With an Infinite Number of Inequality Constraints

We consider a design problem which is subject to a set of uncertain constraints. Suppose that given functions $f_i : \Theta \times \mathcal{W} \rightarrow \mathbb{R}$, $i = 1, \ldots, n$ one desires to obtain $\theta \in \Theta$ such that the (infinite) set of constraints

$$f_i(\theta, w) \leq 0, \; i = 1, \ldots, n, \text{ for all } w \in \mathcal{W}$$

is satisfied. If the problem is feasible, we may choose a specific solution optimizing a given criterion, and this requires to solve a semi-infinite optimization problem of the form

$$\min_{\theta\in\Theta} J(\theta) \text{ subject to } f_i(\theta, w) \leq 0, \; i = 1, \ldots, n, \text{ for all } w \in \mathcal{W}.$$

Then, design with semi-infinite constraints falls into the class of problems considered in this paper, provided that the function $g : \Theta \times \mathcal{W} \rightarrow \{0,1\}$ is defined as

$$g(\theta, w) := \begin{cases} 0 & \text{if } f_i(\theta, w) \leq 0, \; i = 1, \ldots, n, \\ 1 & \text{otherwise.} \end{cases}$$

■

B. Example 2: Min-Max Problem With Uncertainty

In this example, we show that the problem (2) is closely related to the min-max problem, which is studied, for example, in the theory of differential games [7]. Consider a measurable function $\tilde{J} : \Xi \times \mathcal{W} \rightarrow (-\infty, \infty)$, where Ξ is a bounded set, and the min-max design problem

$$\min_{\zeta\in\Xi} \sup_{w\in\mathcal{W}} \tilde{J}(\zeta, w). \tag{3}$$

Taking $\Theta = \{(\zeta, \gamma) : \zeta \in \Xi \text{ and } \gamma \in (-\infty, \infty)\}$, the binary function $g : \Theta \times \mathcal{W} \rightarrow \{0,1\}$ is defined as follows: given $\theta = (\zeta, \gamma) \in \Theta$ and $w \in \mathcal{W}$, we have

$$g(\theta, w) := \begin{cases} 0 & \text{if } \tilde{J}(\zeta, w) \leq \gamma, \\ 1 & \text{otherwise.} \end{cases}$$

Given $\theta = (\zeta, \gamma)$, $J(\theta)$ is defined as $J(\theta) := \gamma$. Then, the min-max problem can be immediately reformulated as

$$\min_{\theta\in\Theta} J(\theta) \text{ subject to } g(\theta, w) = 0, \quad \text{for all } w \in \mathcal{W}. \tag{4}$$

We observe that in this case the feasibility set is always non empty because $\tilde{J} : \Xi \times \mathcal{W} \rightarrow (-\infty, \infty)$ is a bounded function. ■

The robust design problems formulated in this paper are generally very difficult to solve because in many situations the equality $g(\theta, w) = 0$ is *not* a convex constraint on the decision variable θ. Moreover, the set $\mathcal{W}$ has infinite cardinality. One way of circumventing this issue consists in the use of randomization, see [32], [37]. This paper is focused on computing manageable bounds on the sample size to guarantee that the proposed randomized scheme yields an appropriate (in a probabilistic sense) solution to the semi-infinite design problems, see [1]–[3] for preliminary results along this direction.

The main contribution of this paper is to show that the sample size bounds which appeared in the control literature can be reduced by several order of magnitude for "reasonable" values of confidence δ and accuracy ϵ using "one-sided inequalities" (instead of the more classical "two-sided" inequalities). Furthermore, we present a new result that applies to feasibility and optimization problems involving Boolean functions. Finally, we introduce a randomized algorithm which makes use of the explicit bounds derived in the paper. The algorithm is then used for an application example dealing with the controller design of the lateral motion of an aircraft.

II. A Randomized Strategy

We assume that a probability measure $\text{Pr}_{\mathcal{W}}$ over the sample space $\mathcal{W}$ is given; L, M and N represent positive integers. Then, given $\mathcal{W}$, a collection of N independent identically distributed (i.i.d.) samples $\text{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ drawn from $\mathcal{W}$ is said to belong to the Cartesian product $\mathcal{W}^N = \mathcal{W} \times \cdots \times \mathcal{W}$ (N times). Moreover, if the collection w of N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ is generated from $\mathcal{W}$ according to the probability measure $\Pr_{\mathcal{W}}$, then the *multisample* w is drawn according to the probability measure $\Pr_{\mathcal{W}^N}$. The *accuracy*, *confidence* and constraint level (or *level*) are denoted by $\epsilon \in (0,1)$, $\delta \in (0,1)$ and $\rho \in [0,1)$, respectively. For $x \in \mathbb{R}$, $x > 0$, $\lceil x \rceil$ denotes the minimum integer greater than or equal to x, $\ln(\cdot)$ is the natural logarithm and e is the Euler number.

A. Probability of Violation and Empirical Mean

Given $\theta \in \Theta$, there might be a fraction of the elements of $\mathcal{W}$ for which the constraint $g(\theta, w) = 0$ is not satisfied. This concept is rigorously formalized by means of the notion of "probability of violation" which is now introduced.

Definition 1: (Probability of Violation): Consider a probability measure $\Pr_{\mathcal{W}}$ over $\mathcal{W}$ and let $\theta \in \Theta$ be given. The probability of violation of θ for the function $g : \Theta \times \mathcal{W} \to \{0,1\}$ is defined as

$$E_g(\theta) := \Pr_{\mathcal{W}} \{w \in \mathcal{W} : g(\theta, w) = 1\}.$$

Given $\theta \in \Theta$, it is generally difficult to obtain the exact value of the probability of violation $E_g(\theta)$ since this requires the solution of a multiple integral. However, we can approximate its value using the concept of empirical mean. For given $\theta \in \Theta$, the empirical mean of $g(\theta, w)$ with respect to the multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ is defined as

$$\hat{E}_g(\theta, \mathrm{w}) := \frac{1}{N} \sum_{i=1}^{N} g(\theta, w^{(i)}). \tag{5}$$

Clearly, the empirical mean $\hat{E}_g(\theta, \mathrm{w})$ is a random variable. Since $g(\cdot,\cdot)$ is a binary function, $\hat{E}_g(\theta, \mathrm{w})$ is always within the closed interval [0, 1]. As discussed below, the empirical mean can be used to formulate a randomized approach for the solution of the general semi-infinite problems considered in this paper.

B. Randomized Feasibility and Optimization Problems

Suppose that a probability measure $\Pr_{\mathcal{W}}$ over the set $\mathcal{W}$ and the level $\rho \in [0,1)$ are given. Consider the following *randomized strategy*:

i) Draw N i.i.d. samples $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ according to the probability $\Pr_{\mathcal{W}}$.

ii) Find (if possible) a feasible solution $\theta \in \Theta$ of the constraint

$$\hat{E}_g(\theta, \mathrm{w}) \le \rho. \tag{6}$$

iii) If a feasible solution exists, solve the optimization problem

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } \hat{E}_g(\theta, \mathrm{w}) \le \rho. \tag{7}$$

In this paper, the problems (6) and (7) are denoted as *randomized feasibility* and *randomized optimization*, respectively. We note that considering a level ρ larger than zero broadens the class of problems that can be addressed by the proposed methodology. Furthermore, taking $\rho > 0$ allows one to deal with probabilistic (soft) constraints of the kind $\hat{E}_g(\theta, \mathrm{w}) \le \rho$, while deterministic (hard) constraints can be incorporated in the definition of the set Θ of decision variables. We observe that problem (7) is computationally difficult and its solution requires the development of specific techniques and algorithms (see Section VII for further details).

We remark that related randomized strategies can be found in the literature. For example, empirical mean minimization techniques are proposed in [39], min-max problems are presented in [18] and bootstrap learning methods are introduced in [23]. In these references, the main emphasis is on deriving explicit bounds on the required number of samples randomly drawn from $\mathcal{W}$ to guarantee that the obtained solution satisfies some probabilistic accuracy and confidence specifications. All these methods, however, are normally based also on the randomization of the design parameter set Θ, which implies that a finite family approach is followed. That is, only a subset of finite cardinality of Θ is considered. Finally, we recall that the choice of the appropriate probability measure has been studied for example in [8] and [9].

In the next section, we introduce different probabilistic notions that are required to assess the probabilistic properties of the solution obtained by the proposed randomized strategy.

III. Estimating the Probability of Failure

We address the problem of obtaining an explicit bound on the sample size to guarantee that the empirical mean is within a pre-specified accuracy $\epsilon \in (0,1)$ from the probability of violation with high confidence $1 - \delta$, $\delta \in (0,1)$. The Hoeffding inequalities (see for example [19], [32]) characterize how the empirical mean approximates, from a probabilistic point of view, the exact value of $E_g(\theta)$. Suppose that $\epsilon \in (0,1)$ is given. Then, from the Hoeffding inequalities we conclude that

$$\Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N : |E_g(\theta) - \hat{E}_g(\theta, \mathrm{w})| \ge \epsilon\} \le 2e^{-2N\epsilon^2}. \tag{8}$$

Thus, in order to guarantee confidence $1 - \delta$ it suffices to take N such that $2e^{-2N\epsilon^2} \le \delta$ is satisfied. That is, we obtain the (additive) Chernoff bound [16]

$$N \ge \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}.$$

Unfortunately, this inequality is not applicable to the randomized framework stated in (6) and (7). In fact, the Chernoff bound is valid only for *fixed* θ. In the randomized strategy previously described, the parameter θ is a design variable varying in the set Θ. Motivated by this discussion, we now introduce the definition of probability of two-sided failure.

Definition 2: (Probability of Two-Sided Failure): Given N, $\epsilon \in (0,1)$ and $g : \Theta \times \mathcal{W} \to \{0,1\}$, the probability of two-sided failure, denoted by $q_g(N, \epsilon)$ is defined as

$$q_g(N, \epsilon) := \Pr_{\mathcal{W}^N}\left\{\mathrm{w} \in \mathcal{W}^N : \sup_{\theta \in \Theta} |E_g(\theta) - \hat{E}_g(\theta, \mathrm{w})| > \epsilon\right\}.$$

This definition can be easily explained in words. If N i.i.d. samples are drawn from $\mathcal{W}$ according to the probability $\Pr_{\mathcal{W}}$, the set of elements of Θ having an empirical mean not within ϵ of the exact value $E_g(\theta)$ is empty with probability no smaller than $1 - q_g(N, \epsilon)$. The function $g : \Theta \times \mathcal{W} \to \{0,1\}$ enjoys the Uniform Convergence of Empirical Means property (UCEM property) if $q_g(N, \epsilon) \to 0$ as $N \to \infty$ for each $\epsilon > 0$, see [36], [37]. The probability of two-sided failure is based on the two inequalities $E_g(\theta) - \hat{E}_g(\theta, \mathrm{w}) \leq \epsilon$ and $E_g(\theta) - \hat{E}_g(\theta, \mathrm{w}) \geq -\epsilon$.

In the randomized strategy presented in this paper, we are interested in addressing the following question: after drawing N i.i.d. samples from $\mathcal{W}$, suppose that we find $\hat{\theta} \in \Theta$ such that the corresponding empirical mean is smaller than ρ; then, what is the probability that the difference between $E_g(\hat{\theta})$ and the obtained empirical mean is larger than ϵ? To answer this question, we now introduce the formal definition of probability of one-sided constrained failure.

Definition 3: (Probability of One-Sided Constrained Failure): Given N, $\epsilon \in (0,1)$, $\rho \in [0,1)$ and $g : \Theta \times \mathcal{W} \to \{0,1\}$, the probability of one-sided constrained failure, denoted by $p_g(N, \epsilon, \rho)$ is defined as

$$p_g(N, \epsilon, \rho) := \mathrm{Pr}_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N : \text{There exists } \theta \in \Theta \text{ such that } \hat{E}_g(\theta, \mathrm{w}) \leq \rho \text{ and } E_g(\theta) > \hat{E}_g(\theta, \mathrm{w}) + \epsilon\}.$$

Finally we introduce the probability of relative difference failure which is instrumental to prove some of the main results of the paper.

Definition 4: (Probability of Relative Difference Failure): Given N, $\epsilon \in (0,1)$ and $g : \Theta \times \mathcal{W} \to \{0,1\}$, the probability of relative difference failure, denoted by $r_g(N, \epsilon)$, is defined as

$$r_g(N, \epsilon) := \mathrm{Pr}_{\mathcal{W}^N}\left\{\mathrm{w} \in \mathcal{W}^N : \sup_{\theta \in \Theta} \frac{E_g(\theta) - \hat{E}_g(\theta, \mathrm{w})}{\sqrt{E_g(\theta)}} > \epsilon\right\}.$$

The previous definitions introduce two (one-sided) notions of probability of failure. As it will be shown in this paper, when a constraint on the empirical mean is given, one can obtain sample size bounds which are several order of magnitude smaller than those derived from the probability of two-sided failure. The next result shows how to bound the probability of one-sided constrained failure by means of the probability of two-sided failure and the probability of relative difference failure.

Theorem 1: The inequalities

$$p_g(N, \epsilon, \rho) \leq q_g(N, \epsilon) \quad (9)$$

$$p_g(N, \epsilon, \rho) \leq r_g\left(N, \frac{\epsilon}{\sqrt{\epsilon + \rho}}\right) \quad (10)$$

hold for every N, $\epsilon \in (0,1)$, $\rho \in [0,1)$ and $g : \Theta \times \mathcal{W} \to \{0,1\}$.

Proof: See Appendix 1.

IV. Some One-Sided and Two-Sided Results From Statistical Learning Theory

All the results in this section are available in the literature on statistical learning theory [34], [37]. They are stated here for completeness and because they are instrumental to subsequent results presented in Sections V and VI.

Formally, let $\mathcal{G}$ denote the family of functions $\{g(\theta, \cdot) : \theta \in \Theta\}$, where $g : \Theta \times \mathcal{W} \to \{0,1\}$. Then, given the multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\} \in \mathcal{W}^N$, the binary vector $(g(\theta, w^{(1)}), \ldots, g(\theta, w^{(N)})) \in \{0,1\}^N$ can attain at most 2^N distinct values when θ varies in Θ. The maximum number of distinct binary vectors (denoted in the following by $\phi_g(\mathrm{w})$) that can be obtained grows with the number of samples N. The next definition introduces in a formal way the notion of growth function (also known as shatter coefficient) [34], [35], [37].

Definition 5: (Growth Function): Given the function $g : \Theta \times \mathcal{W} \to \{0,1\}$ and the multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\} \in \mathcal{W}^N$, $\phi_g(\mathrm{w})$ denotes the number of distinct binary vectors

$$(g(\theta, w^{(1)}), \ldots, g(\theta, w^{(N)})) \in \{0,1\}^N$$

that can be obtained with the different elements of Θ. Then, the growth function $\pi_g(N)$ is defined as

$$\pi_g(N) = \sup_{\mathrm{w} \in \mathcal{W}^N} \phi_g(\mathrm{w}).$$

In words, $\phi_g(\mathrm{w})$ is the cardinality of the set $\{(g(\theta, w^{(1)}), \ldots, g(\theta, w^{(N)})) : \theta \in \Theta\}$ and the growth function is the supremum of $\phi_g(\mathrm{w})$ taken with respect to $\mathrm{w} \in \mathcal{W}^N$.

In the celebrated work of Vapnik and Chervonenkis [35], it has been demonstrated that the growth function can be used to bound the probability of two-sided failure. This is now formally stated.

Theorem 2: Given the function $g : \Theta \times \mathcal{W} \to \{0,1\}$ and $\epsilon \in (0,1)$

$$q_g(N, \epsilon) \leq 4\pi_g(2N)e^{-N\epsilon^2/8}.$$

We are now ready to introduce the notion of VC-dimension, also known as the Vapnik-Chervonenkis dimension, see [34], [37] and the end of this section for additional details.

Definition 6: Given the function $g : \Theta \times \mathcal{W} \to \{0,1\}$, the VC-dimension, denoted as VC_g, is the largest integer N for which the equality $\pi_g(N) = 2^N$ is satisfied.

The VC-dimension establishes the "richness" of a given family $\mathcal{G}$ and it can be used to determine a bound on the growth function by means of the so-called Sauer Lemma, see [28], [37].

Lemma 1: Let the function $g : \Theta \times \mathcal{W} \to \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Then, for every $N \geq d$ we have

$$\pi_g(N) \leq \left(\frac{eN}{d}\right)^d.$$

One of the direct consequences of this result is that, under the assumption of finite VC-dimension, $\pi_g(N)$ is bounded by a polynomial function of N. In turn, this implies that

$$\lim_{N \to \infty} 4\pi_g(2N)e^{-N\epsilon^2/8} = 0.$$

Bearing in mind Theorem 2, we conclude that, if the VC-dimension is bounded, the probability of two-sided failure converges to zero when N tends to infinity. Using Theorem 2 and Lemma 1, the following result is obtained in [37], see Theorems 7.2 and 10.2.

Theorem 3: Let $g : \Theta \times \mathcal{W} \to \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\epsilon \in (0,1)$ and $\delta \in (0,1)$ are given and $N \geq d$. Then

$$q_g(N,\epsilon) \leq 4\left(\frac{2eN}{d}\right)^d e^{-N\epsilon^2/8}.$$

Moreover, $q_g(N,\epsilon) \leq \delta$ provided that

$$N \geq \max\left\{\frac{16}{\epsilon^2}\ln\frac{4}{\delta}, \frac{32d}{\epsilon^2}\ln\frac{32e}{\epsilon^2}\right\}.$$

Theorem 3 is well-known in the control community. However, as discussed in [37], there exist other results in the literature that allow one to reduce by a factor close to 8 the explicit bound on the number of samples. Basically, these results differ from Theorem 2 in the exponent $-N\epsilon^2/8$ which is replaced by less conservative ones. For example, the exponent $-N\epsilon^2$ can be found in [26] and [34]. The result proved in [34] is used in this paper because the exponent $-N\epsilon^2$ is obtained with almost no increase in the other constants appearing in the bound of $q_g(N,\epsilon)$.

The concept of annealed entropy, see Sections 3.13 and 4.4 in [34], is crucial for stating the next result, which is presented in [34], Theorem 4.1.

Theorem 4: Given the function $g : \Theta \times \mathcal{W} \to \{0,1\}$, N and $\epsilon \in (0,1)$

$$q_g(N,\epsilon) < 4e^{\ln \pi_g(2N) - N(\epsilon - 1/N)^2}.$$

Next, for the sake of completeness, we present a corollary which is essentially equivalent to the first claim of Theorem 4.4 of [34].

Corollary 1: Given the function $g : \Theta \times \mathcal{W} \to \{0,1\}$, N and $\epsilon \in (0,1)$

$$q_g(N,\epsilon) < 4e^{2\epsilon}\pi_g(2N)e^{-N\epsilon^2}.$$

Proof: To prove the result, it suffices to consider Theorem 4 and to show a chain of inequalities

$$\begin{aligned} q_g(N,\epsilon) &< 4\pi_g(2N)e^{-(\epsilon-1/N)^2 N} \\ &< 4\pi_g(2N)e^{-\epsilon^2 N + 2\epsilon} \\ &= 4e^{2\epsilon}\pi_g(2N)e^{-N\epsilon^2}. \end{aligned}$$

∎

We remark that a very similar bound $q_g(N,\epsilon) < 6e^{2\epsilon}\pi_g(2N)e^{-N\epsilon^2}$ is shown in [26]. We now state a theorem which shows that the growth function can be used to bound the probability of relative difference failure, see Theorem 4.2 in [34].

Theorem 5: Given the function $g : \Theta \times \mathcal{W} \to \{0,1\}$, N and $\epsilon \in (0,1)$

$$r_g(N,\epsilon) < 4\pi_g(2N)e^{-N\epsilon^2/4}.$$

Combining Lemma 1, Corollary 1 and Theorem 5 we obtain a corollary which is now stated.

Corollary 2: Let $g : \Theta \times \mathcal{W} \to \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\epsilon \in (0,1)$ is given, and $N \geq d$. Then

$$q_g(N,\epsilon) < 4e^{2\epsilon}\left(\frac{2eN}{d}\right)^d e^{-N\epsilon^2} \tag{11}$$

$$r_g(N,\epsilon) < 4\left(\frac{2eN}{d}\right)^d e^{-N\epsilon^2/4}. \tag{12}$$

We notice that the first claim of this corollary is less conservative than Theorem 3 because the exponent $-N\epsilon^2/8$ is replaced with $-N\epsilon^2$ even though the factor 4 is now replaced by $4e^{2\epsilon}$.

Finally, we now recall a bound on the VC-dimension stated in [37]. This bound has been extensively used to compute the VC-dimension when the constraints of the problem can be written as Booleans functions (see definition below). This bound is useful in many classical control problems [39], [40], including the well-known static output feedback, see also the example in Section VIII.

Definition 7: ((α,k)-Boolean Function): The function $g : \Theta \times \mathcal{W} \to \{0,1\}$ is an (α,k)-Boolean function if for fixed w it can be written as an expression consisting of Boolean operators involving k polynomials

$$\beta_1(\theta), \beta_2(\theta), \ldots, \beta_k(\theta)$$

in the components $\theta_i, i = 1,\ldots,n_\theta$, and the degree with respect to θ_i of all these polynomials is no larger than α.

Lemma 2: Suppose that $g : \Theta \times \mathcal{W} \to \{0,1\}$ is an (α,k)-Boolean function. Then

$$\mathrm{VC}_g \leq 2n_\theta \log_2(4e\alpha k).$$

This result, which was stated in [37], relies on a similar bound given in [22].

V. Explicit Sample Size Bounds for the Probability of Failure

In this section, we focus on the problem of obtaining explicit bounds on the number of samples required to guarantee a pre-specified confidence δ. More specifically, given $\delta \in (0,1)$, and real constants a, b, and c, the results of this section allow one to obtain $N \geq d$ satisfying the inequality

$$a\left(\frac{ceN}{d}\right)^d e^{-bN} < \delta. \tag{13}$$

This result, which is proved in Appendix 2, is now formally stated.

Theorem 6: Suppose that $a \geq 1$, $b \in (0,1]$, $c \geq 1$, $d \geq 1$ and $\delta \in (0,1)$ are given. Then, $N \geq d$ and

$$a\left(\frac{ceN}{d}\right)^d e^{-bN} < \delta \tag{14}$$

provided that

$$N \geq \inf_{\mu > 1} \frac{\mu}{b(\mu - 1)}\left(\ln\frac{a}{\delta} + d\ln\frac{c\mu}{b}\right).$$

As an example of an application of Theorem 6, we obtain an alternative explicit bound to that presented in the second claim of Theorem 3. From the first claim of this result, to achieve a probability of two-sided failure smaller than δ, it suffices to choose N such that

$$4\left(\frac{2eN}{d}\right)^d e^{-N\epsilon^2/8} < \delta.$$

Taking $a = 4$, $b = \epsilon^2/8$, $c = 2$ and applying Theorem 6, we easily obtain the bound

$$N \geq \inf_{\mu>1} \frac{8}{\epsilon^2}\left(\frac{\mu}{\mu-1}\right)\left(\ln\frac{4}{\delta} + d\ln\frac{16\mu}{\epsilon^2}\right).$$

Clearly, a conservative bound may be computed if, instead of evaluating the infimum with respect to $\mu > 1$, a suboptimal value of μ is chosen. For example, setting μ equal to $2e$, we conclude that it suffices to take

$$N \geq \frac{9.81}{\epsilon^2}\left(\ln\frac{4}{\delta} + d\ln\frac{32e}{\epsilon^2}\right).$$

We remark that this bound is less conservative than that stated in Theorem 3. In the next corollary, we show that further improvements with respect to Theorem 3 can be obtained using Theorem 6 and Corollary 2.

Corollary 3: Let $g : \Theta \times \mathcal{W} \to \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\epsilon \in (0,1)$ and $\delta \in (0,1)$ are given. Then, the probability of two-sided failure $q_g(N,\epsilon)$ is smaller than δ if

$$N \geq \frac{1.2}{\epsilon^2}\left(\ln\frac{4e^{2\epsilon}}{\delta} + d\ln\frac{12}{\epsilon^2}\right).$$

Proof: From the first claim of Corollary 2 we conclude that it suffices to choose N such that

$$4e^{2\epsilon}\left(\frac{2eN}{d}\right)^d e^{-N\epsilon^2} < \delta.$$

Taking $a = 4e^{2\epsilon}$, $b = \epsilon^2$, $c = 2$ and $\mu = 6$, the result follows immediately from Theorem 6. ■

We recall that Theorem 1 states that $p_g(N,\epsilon,\rho) \leq q_g(N,\epsilon)$ for all N, $\epsilon \in (0,1)$ and $\rho \in [0,1)$. This means that Corollary 3 can be used to bound the probability of one-sided constrained failure. Although Corollary 3 is a clear improvement with respect to Theorem 3, the obtained sample size bound still grows with $1/\epsilon^2 \ln 1/\epsilon^2$. This dependence with respect to ϵ makes the bound of practical interest only for relatively large values of the accuracy parameter $\epsilon \in (0,1)$. Next, we show that, when the level parameter ρ is chosen close to zero, manageable sample size bounds are obtained for reasonable values of accuracy and confidence parameters ϵ and δ. This constitutes a significant computational improvement that reduces by several orders of magnitude the required sample size. Notice that it makes sense to choose ρ close to zero because in most applications one desires to have a small probability of violation.

The next theorem is one of the main contributions of the paper.

Theorem 7: Let $g : \Theta \times \mathcal{W} \to \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\epsilon \in (0,1)$, $\delta \in (0,1)$ and $\rho \in [0,1)$ are given. Then, the probability of one-sided constrained failure $p_g(N,\epsilon,\rho)$ is smaller than δ if

$$N \geq \frac{5(\rho+\epsilon)}{\epsilon^2}\left(\ln\frac{4}{\delta} + d\ln\frac{40(\rho+\epsilon)}{\epsilon^2}\right).$$

Proof: Using Theorem 1, to guarantee $p_g(N,\epsilon,\rho) < \delta$, it suffices to choose N such that $r_g(N,\epsilon/\sqrt{\epsilon+\rho}) < \delta$. From Theorem 5, we conclude that this inequality is satisfied if

$$4\left(\frac{2eN}{d}\right)^d e^{-N\epsilon^2/4(\epsilon+\rho)} < \delta.$$

Next, setting $a = 4$, $c = 2$ and $b = \epsilon^2/4(\epsilon+\rho)$, using Theorem 6 we conclude that it suffices to take

$$N \geq \inf_{\mu>1} \frac{4(\epsilon+\rho)\mu}{\epsilon^2(\mu-1)}\left(\ln\frac{4}{\delta} + d\ln\frac{8(\epsilon+\rho)\mu}{\epsilon^2}\right).$$

Finally, the statement of the theorem follows simply taking $\mu = 5$. ■

In some applications, a reasonable choice of the level parameter is $\rho = \epsilon$. In turn, this means that, if $\hat{\theta}$ satisfies the feasible constraint of the inequality (6) and $p_g(N,\epsilon,\rho) < \delta$, then from the definition of $p_g(N,\epsilon,\rho)$, $E_g(\hat{\theta}) \leq 2\epsilon$ with probability no smaller than $1-\delta$. Taking $\rho = \epsilon$ in Theorem 7 one immediately obtains the bound

$$N \geq \frac{10}{\epsilon}\left(\ln\frac{4}{\delta} + d\ln\frac{80}{\epsilon}\right)$$

which grows with $1/\epsilon \ln 1/\epsilon$. The next corollary provides an explicit bound for the particular case when the level parameter is equal to zero.

Corollary 4: Let $g : \Theta \times \mathcal{W} \to \{0,1\}$ be such that $\mathrm{VC}_g \leq d < \infty$. Suppose that $\epsilon \in (0,1)$ and $\delta \in (0,1)$ are given. Then, the probability of one-sided constrained failure $p_g(N,\epsilon,0)$ is smaller than δ if

$$N \geq \frac{5}{\epsilon}\left(\ln\frac{4}{\delta} + d\ln\frac{40}{\epsilon}\right).$$

It is worth remarking that a similar bound for the particular case $\rho = 0$ can be obtained using Theorem 6 and the notion of "version space" presented in [26]. More generally, we can introduce a parametrization of the form $\rho = \epsilon^{\ell}$, where ℓ is any non negative scalar, which relates ρ and ϵ. With this parametrization the sample size bound grows as $(\epsilon^{\ell-1}+1)/\epsilon \ln (\epsilon^{\ell-1}+1)/\epsilon$.

VI. Explicit Sample Size Bounds for the Zero Level Case

In this section, we present results for the special case $\rho = 0$. In particular, in Subsection VI-A we deal with feasibility and optimization problems with Boolean polynomials constraints and in Subsection VI-B we present a result for convex optimization problems. To this end, we consider the following *zero-level randomized strategy*:

Suppose that a probability measure $\mathrm{Pr}_{\mathcal{W}}$ over the set $\mathcal{W}$ is given.

i) Draw N i.i.d. samples $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ according to the probability $\mathrm{Pr}_{\mathcal{W}}$.

ii) Find (if possible) a feasible solution $\theta \in \Theta$ of the constraint

$$\hat{E}_g(\theta, \mathrm{w}) = 0. \tag{15}$$

iii) If a feasible solution exists, solve the optimization problem

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } \hat{E}_g(\theta, \mathrm{w}) = 0. \tag{16}$$

A. Explicit Sample Size Bounds for Boolean Polynomial Constraints

In this subsection, we consider the zero-level problem when $g : \Theta \times \mathcal{W} \to \{0,1\}$ can be written as an (α, k)-Boolean function (see Definition 7) on the decision variable $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$. We derive a one-sided result that complements the bound presented in Section V, Corollary 4.

Theorem 8: Suppose that $g : \Theta \times \mathcal{W} \to \{0,1\}$ is an (α, k)-Boolean function. Then, given $\epsilon \in (0, 0.14)$ and $\delta \in (0,1)$, if

$$N \geq \bar{N}(\epsilon, \delta) := \frac{4.1}{\epsilon}\left(\ln \frac{21.64}{\delta} + 4.39 n_\theta \log_2\left(\frac{8e\alpha k}{\epsilon}\right)\right)$$

then, with probability no smaller than $1-\delta$, either the optimization problem (16) is infeasible and, hence, also the general optimization problem (2) is infeasible; or, (16) is feasible, and then any feasible solution $\hat{\theta}_N$ satisfies the inequality $E_g(\hat{\theta}_N) \leq \epsilon$.

Proof: The proof is given in Appendix 3 and it relies on the so-called *pack-based strategy* (see [1], [3]).

Note that using the equation (30) in Appendix 3, more general results in which ϵ is not constrained within the open interval $(0, 0.14)$ can be derived. In this sense, the restriction $\epsilon \in (0, 0.14)$ can be relaxed at the expense of obtaining larger constants appearing in the bound $\bar{N}(\epsilon, \delta)$.

We notice that if $\epsilon \in (0, 0.14)$ and $2\epsilon e\alpha k \leq 1$, then

$$\log_2 \frac{8e\alpha k}{\epsilon} \leq \log_2 \frac{4}{\epsilon^2} = 2\log_2 \frac{2}{\epsilon}.$$

Therefore, in this case the bound provided in Theorem 8 yields

$$N \geq \frac{4.1}{\epsilon}\left(\ln \frac{21.64}{\delta} + 8.78 n_\theta \log_2 \frac{2}{\epsilon}\right).$$

That is, if ϵ is sufficiently small, we obtain an explicit bound that only depends on ϵ, n_θ and δ.

B. The Scenario Approach

In this subsection, we study the so-called scenario approach for robust control introduced in [11], [12], see also [15] for recent results in this area. We address the semi-infinite optimization problem

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } g(\theta, w) = 0, \quad \text{for all } w \in \mathcal{W} \tag{17}$$

for the particular case in which $J(\theta) = c^T\theta$, the constraint $g(\theta, w) = 0$ is convex in θ for all $w \in W$, the solution of (17) is unique[1] and the level parameter ρ is equal to zero. The first two assumptions are now stated precisely.

[1]We remark that this uniqueness assumption can be relaxed in most cases, as shown in Appendix 1 of [12].

Assumption 1: (Convexity): Let $\Theta \subset \mathbb{R}^{n_\theta}$ be a convex and closed set, and let $\mathcal{W} \subseteq \mathbb{R}^{n_w}$. We assume that

$$J(\theta) := c^T\theta \quad \text{and}$$
$$g(\theta, w) := \begin{cases} 0 & \text{if } f(\theta, w) \leq 0, \\ 1 & \text{otherwise} \end{cases}$$

where $f(\theta, w) : \Theta \times \mathcal{W} \to [-\infty, \infty]$ is convex in θ for any fixed value of $w \in \mathcal{W}$.

Assumption 2: (Uniqueness): The optimization problem (16) is either infeasible, or, if feasible, it attains a unique optimal solution for all possible multisample extractions $\{w^{(1)}, \ldots, w^{(N)}\}$.

We state here a result that was first presented in [12].

Corollary 5: Let Assumptions 1 and 2 hold. Suppose that N, $\epsilon \in (0,1)$ and $\delta \in (0,1)$ satisfy the following inequality:

$$N \geq \frac{2}{\epsilon} \ln \frac{1}{\delta} + 2n_\theta + \frac{2n_\theta}{\epsilon} \ln \frac{2}{\epsilon}. \tag{18}$$

Then, with probability no smaller than $1-\delta$, either the optimization problem (16) is infeasible and, hence, also the semi-infinite convex optimization problem (17) is infeasible; or, (16) is feasible, and then its optimal solution $\hat{\theta}_N$ satisfies the inequality $E_g(\hat{\theta}_N) \leq \epsilon$.

The following result is based on the pack-based formulation presented in Appendix 3 and on the proof of Theorem 1 of [12].

Corollary 6: Let Assumptions 1 and 2 be satisfied. Then, given $\epsilon \in (0,1)$ and $\delta \in (0,1)$, Corollary 5 holds for

$$N \geq \frac{2}{\epsilon} \ln \frac{1}{2\delta} + 2n_\theta + \frac{2n_\theta}{\epsilon} \ln 4. \tag{19}$$

Proof: See [1] for a proof.

The contribution of Corollary 6 is to demonstrate that the term $\frac{2n_\theta}{\epsilon} \ln \frac{2}{\epsilon}$ in (18) can be replaced with $\frac{2n_\theta}{\epsilon} \ln 4$. As a direct consequence, we infer that the required sample size is inversely proportional to the accuracy parameter for fixed confidence. We notice that the constants appearing in the bound may be slightly improved.

VII. A Randomized Algorithm for Semi-Infinite Problems

We now present a randomized algorithm for the solution of the semi-infinite problems stated in (1) and (2). This algorithm has some similarities with that presented in [23] for finite families of controllers. Assuming that the function $g : \Theta \times \mathcal{W} \to \{0,1\}$ has finite VC-dimension $\mathrm{VC}_g \leq d$, we provide a strategy that allows one to circumvent the potential conservativeness of the sample size bounds previously obtained. This is accomplished by means of a sequence of optimization problems of increasing complexity where the number of iterations is bounded by the termination parameter k_t. As it can be seen in the algorithm below, at each iteration, two sets of samples of cardinality N_k and M_k are generated. The first set, consisting of samples $\{w^{(1)}, \ldots, w^{(N_k)}\}$, is used to obtain a candidate probabilistic solution $\hat{\theta}_{N_k}$ to the optimization problem. The performance of this candidate solution is then tested using the validations samples $\{v^{(1)}, \ldots, v^{(M_k)}\}$.

The cardinality of these sets grows at each iteration k as

$$N_k = \left\lceil \beta_p^k \left(\frac{\rho+\epsilon}{\epsilon^2}\right)\right\rceil,$$
$$M_k = \left\lceil 2\beta_v^k \left(\frac{\rho+\epsilon}{\epsilon^2}\right) \ln \frac{2k_t}{\delta}\right\rceil.$$

The constants β_p and β_v are chosen such that $N_{k_t} = \hat{N}_t$ and $M_{k_t} = \max\left\{\hat{N}_t, \lceil 2\left(\frac{\rho+\epsilon}{\epsilon^2}\right)\ln\frac{2k_t}{\delta}\rceil\right\}$, where $\hat{N}_t$ is the sample size bound obtained using the results previously presented in this paper (see Theorem 7). The parameter k_t limits the maximal number of iterations of the algorithm and allows the user to control the sample size of the first iteration of the algorithm. The parameter k_t may be chosen in the interval

$$k_t \in [k_t^-, k_t^+] \quad \text{where}$$
$$k_t^- = \left\lceil \frac{\ln(\hat{N}_t\epsilon^2) - \ln(\rho+\epsilon)}{\ln 2}\right\rceil \quad \text{and} \quad k_t^+ = 2k_t^-.$$

If k_t is equal to k_t^-, then β_p is close to 2, N_k approximately doubles at each iteration and the sample size of the first iteration is close to $\frac{2(\rho+\epsilon)}{\epsilon^2}$.

The main idea of the algorithm is that if the candidate solution $\hat{\theta}_{N_k}$ satisfies

$$\frac{1}{M_k}\sum_{i=1}^{M_k} g(\hat{\theta}_{N_k}, v^{(i)}) \le \rho + (1-\beta_v^{-k/2})\epsilon$$

at iteration k, then it can be classified as a probabilistic solution with confidence δ, accuracy ϵ and level ρ and no further iterations of the algorithm are needed. This means that the algorithm may find a solution using a number of samples much smaller than $\hat{N}_t$. The algorithm is now formally stated.

A. Randomized Algorithm

(i) Set ϵ, δ and ρ equal to the desired levels. Choose the integer $k_t \ge 1$. Set $k=0$ and

$$\hat{N}_t = \left\lceil \frac{5(\rho+\epsilon)}{\epsilon^2}\left(\ln\frac{8}{\delta} + d\ln\frac{40(\rho+\epsilon)}{\epsilon^2}\right)\right\rceil.$$

(ii) Set

$$\beta_p = \left(\frac{\hat{N}_t\epsilon^2}{\rho+\epsilon}\right)^{1/k_t},$$
$$\beta_v = \max\left\{1, \beta_p\left(2\ln\frac{2k_t}{\delta}\right)^{-1/k_t}\right\}.$$

(iii) If $k \ge k_t$ then Exit. Else, set $k = k+1$ and $N_k = \lceil \beta_p^k \left(\frac{\rho+\epsilon}{\epsilon^2}\right)\rceil$.

(iv) Draw N_k i.i.d. samples $\{w^{(1)},\dots,w^{(N_k)}\}$ according to the probability $\mathrm{Pr}_{\mathcal{W}}$.

(v) Compute (if possible) a feasible solution $\theta \in \Theta$ of the constraint

$$\frac{1}{N_k}\sum_{i=1}^{N_k} g(\theta, w^{(i)}) \le \rho. \qquad (20)$$

(vi) If no feasible solution to this problem is found, then Goto (iii). Else, compute a (suboptimal) solution $\hat{\theta}_{N_k}$ to the optimization problem

$$\min_{\theta\in\Theta} J(\theta) \text{ subject to } \frac{1}{N_k}\sum_{i=1}^{N_k} g(\theta, w^{(i)}) \le \rho. \qquad (21)$$

(vii) If $k = k_t$ then $\hat{\theta}_{N_k}$ is a probabilistic feasible solution with confidence δ, accuracy ϵ and level ρ. Exit.

(viii) Draw $M_k = \lceil 2\beta_v^k\left(\frac{\rho+\epsilon}{\epsilon^2}\right)\ln\frac{2k_t}{\delta}\rceil$ i.i.d. validation samples $\{v^{(1)},\dots,v^{(M_k)}\}$ according to the probability $\mathrm{Pr}_{\mathcal{W}}$.

(ix) If

$$\frac{1}{M_k}\sum_{i=1}^{M_k} g(\hat{\theta}_{N_k}, v^{(i)}) \le \rho + (1-\beta_v^{-k/2})\epsilon$$

then $\hat{\theta}_{N_k}$ is a probabilistic feasible solution with confidence δ, accuracy ϵ and level ρ. Exit. Else, Goto step (iii).

Remark 1: (Optimization): We notice that in the absence of a convexity assumption, (21) requires the solution of a non convex optimization problem. Previous work [18], [38] focused on the use of randomization to obtain a finite set of random samples from the design parameter set Θ. In this way, the so-called near minima are obtained in polynomial time. However, for controller design, as discussed in [23], a simple Monte Carlo scheme of optimization can be very misleading because the obtained empirical minimum may be much larger than the true one with probability practically equal to one. We therefore propose to use a suitable local search algorithm to solve the (possible) non convex problem (21). Contrary to aforementioned works, this approach is no longer based on a finite family optimization method.

Property 1: (Validation): The proposed algorithm guarantees that, if at iteration k, $\hat{\theta}_{N_k}$ is classified as a probabilistic feasible solution, then $\hat{\theta}_{N_k}$ satisfies $E_g(\hat{\theta}_{N_k}) \le \rho + \epsilon$ with probability no smaller than $1-\delta$.

Proof: The proof relies on the sample size bound presented in Theorem 7. In fact, the proposed value for $\hat{N}_t$ guarantees that $p_g(\hat{N}_t, \epsilon, \rho) \le \delta/2$. A detailed proof of the property is given in Appendix 5.

It is clear that if the original problem (1) is feasible then the problem (20) is also feasible for every iteration of the algorithm. However, if the original problem is not feasible, then the algorithm may exit without providing a feasible probabilistic solution, see the following property.

Property 2: (Feasibility): If the algorithm terminates without providing a probabilistic feasible solution then, with probability no smaller than $1-\delta$, there is no $\theta \in \Theta$ such that $E_g(\theta) \le \rho - \mu$, where

$$\mu := \sqrt{\frac{1}{2\hat{N}_t}\ln\frac{2}{\delta}}.$$

Proof: See Appendix 6.

VIII. Application Example

We consider a multivariable example given in [6] and [32] (see also the original paper [33] for a slightly different model and set of data) which studies the design of a controller for the lateral motion of an aircraft. The state space equation consists of four states and two inputs and is given by

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & L_p & L_\beta & L_r \\ \frac{g}{V} & 0 & Y_\beta & -1 \\ N_{\dot{\beta}} & N_p & N_\beta + N_{\dot{\beta}} Y_\beta & N_r - N_{\dot{\beta}} \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 0 & -3.91 \\ 0.035 & 0 \\ -2.53 & 0.31 \end{bmatrix} u$$

where x_1 is the bank angle, x_2 its derivative, x_3 is the sideslip angle, x_4 the yaw rate, u_1 the rudder deflection and u_2 the aileron deflection.

The following nominal values for the nine aircraft parameters entering into the state matrix are taken: $L_p = -2.93$, $L_\beta = -4.75$, $L_r = 0.78$, $g/V = 0.086$, $Y_\beta = -0.11$, $N_{\dot{\beta}} = 0.1$, $N_p = -0.042$, $N_\beta = 2.601$ and $N_r = -0.29$. Then, each nominal parameter is perturbed by a relative uncertainty which is taken to be equal to 15%; e.g., the parameter L_p is bounded in the interval $[-3.3695, \ -2.4905]$. That is, we consider the uncertain system

$$\dot{x}(t) = A(w)x(t) + Bu(t)$$

where $w = [L_p, \ L_\beta, \ L_r, \ g/V, \ Y_\beta, \ N_{\dot{\beta}}, \ N_p, \ N_\beta, \ N_r]^T$ is a random vector with uniform probability density function. The set $\mathcal{W}$ is an hyper-rectangle which is defined accordingly.

We consider the problem of obtaining a state feedback $u(t) = Kx(t)$ such that the closed loop system $\dot{x}(t) = (A(w) + BK)x(t)$ is Hurwitz for every value of w. Moreover, as in [6], we assume a constraint on the magnitude of the entries of the gain matrix K. That is

$$-\bar{K} \preceq K \preceq \bar{K}$$

where the symbol $\preceq$ stands for entry-wise inequality and

$$\bar{K} = \begin{bmatrix} 5 & 0.5 & 5 & 5 \\ 5 & 2 & 20 & 1 \end{bmatrix}.$$

In particular, we look for the optimal value of K that maximizes the convergence rate γ to the origin. Then, the decision parameter θ consists of K and γ, i.e., $\theta := (K, \gamma)$. The design set Θ is given by

$$\Theta := \{\theta = (K, \gamma) : -\bar{K} \preceq K \preceq \bar{K}, \ \gamma \in \mathbb{R}\}.$$

The function $g : \Theta \times \mathcal{W} \to \{0, 1\}$ is defined as follows: given $\theta = (K, \gamma)$ and $w \in \mathcal{W}$

$$g(\theta, w) := \begin{cases} 0 & \text{if } A(w) + \gamma \mathrm{I} + BK \text{ is Hurwitz} \\ 1 & \text{otherwise.} \end{cases}$$

Summing up, the non convex semi-infinite optimization problem under consideration is

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } g(\theta, w) = 0, \quad \text{for all } w \in \mathcal{W},$$

where $J(\theta) := -\gamma$. Since the condition on Hurwitz stability can be rewritten as a Boolean expression consisting of polynomials, invoking Lemma 2 we compute a bound on the VC-dimension obtaining $\mathrm{VC}_g \leq d = 200$ (see [38] for further details).

In this example, we first set confidence δ, accuracy ϵ and level ρ equal to 10^{-6}, 10^{-2} and 0 respectively. In this case, using Theorem 7 we obtain a sample size $N \geq 835,176$; since ρ is equal to zero, we also compute the sample size bound $N \geq 310,739$ using Theorem 8. Clearly, these are major improvements comparing to the sample size given by Theorem 3, which is $N \geq 8.75 \cdot 10^8$.

To circumvent the potential conservativeness of the sample size bounds for this example, we run the proposed randomized algorithm with the exiting parameter k_t equal to 20. The algorithm provided a probabilistic solution at iteration 5. The sample sizes N_k and M_k at the last iteration were equal to 957 and 13,761 respectively. The obtained controller is given by

$$K = \begin{bmatrix} 1.9043 & 0.5000 & -5.0000 & 2.8951 \\ 5.0000 & 1.5080 & 4.4829 & -1.0000 \end{bmatrix}$$

and the corresponding value for γ is 3.93. The results provided in this paper guarantee that, with confidence $1 - \delta$, for the obtained controller 99% of the uncertain plants (notice that $1 - \rho - \epsilon = 0.99$) have a rate of convergence greater or equal to 3.93. We notice that some of the controller gains are equal to the entries of the bounding matrix $\bar{K}$, which shows that the constraints of the problem take effects into the obtained solution.

Once the controller was computed, we formulated a sufficient condition for robust quadratic stability, obtaining a deterministic worst case value for the convergence rate γ. Then, we obtained the generalized eigenvalue problem

$$\max_{\gamma, P > 0} \gamma \text{ subject to}$$
$$P(A(w) + \gamma \mathrm{I} + BK) + (A(w) + \gamma \mathrm{I} + BK)^T P < 0, \quad \text{for all } w \in \mathcal{W}.$$

Taking into account that w enters affinely in the matrix $A(w)$, we reformulate this as a generalized eigenvalue problem subject to 512 constraints, i.e., the number of vertices of the hyper-rectangle $\mathcal{W}$ (this number of constraints can be reduced using the results of [4]). By means of the LMI-toolbox for Matlab, the matrix

$$P = \begin{bmatrix} 0.1164 & 0.0202 & -0.0822 & 0.0297 \\ 0.0202 & 0.0048 & 0.0105 & 0.0000 \\ -0.0822 & 0.0105 & 0.9472 & -0.2042 \\ 0.0297 & 0.0000 & -0.2042 & 0.0475 \end{bmatrix}$$

and the value $\gamma = 3.44$ have been computed. This guarantees that the controller K derived with the proposed randomized strategy gives a closed loop uncertain system which is robustly stable for all $w \in \mathcal{W}$. Fig. 1 shows the resulting closed loop eigenvalues of the system for 500 elements randomly drawn from $\mathcal{W}$.

We run also the algorithm for $\rho = 0.05$, $\epsilon = 0.01$, $\delta = 10^{-6}$ and $k_t = 20$. In this case we obtained a controller that, with probability no smaller than $1 - \delta$, guarantees that 94% of the uncertain plants (notice that $1 - \rho - \epsilon = 0.94$) have a rate of convergence greater or equal to 4.0. The deterministic sufficient condition provided a worst-case rate of convergence equal to

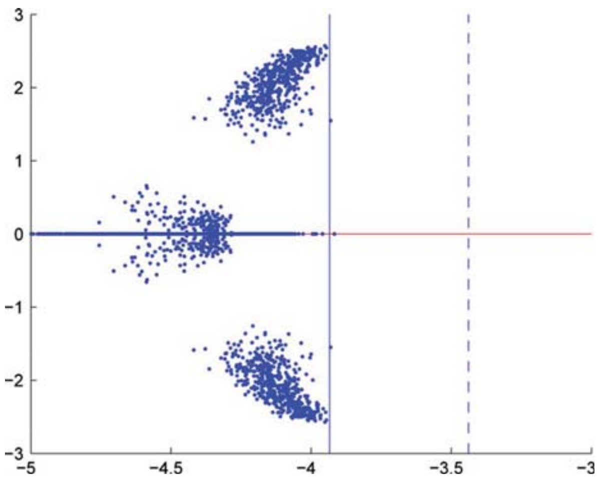

Fig. 1. Closed loop eigenvalues for $\rho = 0$, $\epsilon = 0.01$ and $\delta = 10^{-6}$.

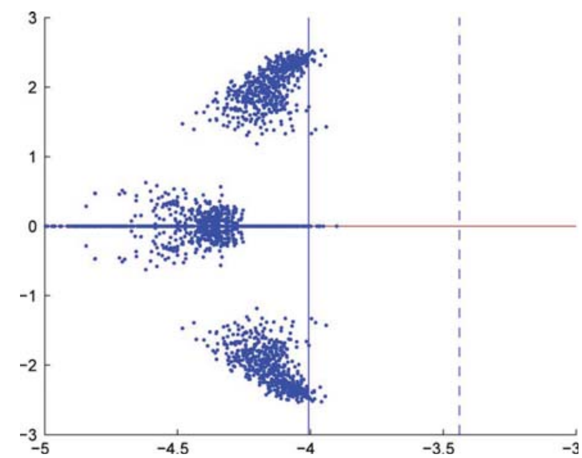

Fig. 2. Closed loop eigenvalues for $\rho = 0.05$, $\epsilon = 0.01$ and $\delta = 10^{-6}$.

3.43. Fig. 2 shows the resulting closed loop eigenvalues of the system for 500 elements randomly drawn from $\mathcal{W}$.

IX. Conclusion

In this paper, we proposed a randomized strategy to address a general class of semi-infinite feasibility and optimization problems. We have shown the usefulness of constrained one-sided results from statistical learning theory for successfully addressing various uncertain systems and control problems. We have demonstrated that the obtained bounds are of several order of magnitude smaller than those currently utilized in the control community literature for "reasonable" values of probabilistic confidence δ and accuracy ϵ. In particular, we have shown that the number of required samples grows with the accuracy parameter ϵ as $1/\epsilon \ln 1/\epsilon$, and this is a significant improvement when compared to the existing bounds which depend on $1/\epsilon^2 \ln 1/\epsilon^2$.

Appendix I
Proof of Theorem 1

First we prove the inequality (9). Clearly, $p_g(N,\epsilon,\rho) \leq p_g(N,\epsilon,1)$, for every N, $\epsilon \in (0,1)$ and $\rho \in [0,1)$. Since $\hat{E}_g(\theta,\mathrm{w}) \leq 1$, for all $\theta \in \Theta$ and $\mathrm{w} \in \mathcal{W}^N$, it results that

$$\begin{aligned} p_g(N,\epsilon,1) = &\Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N : \text{There exists } \theta \in \Theta \\ &\text{such that } E_g(\theta) > \hat{E}_g(\theta,\mathrm{w}) + \epsilon\} \leq \\ \Pr_{\mathcal{W}^N}&\left\{\mathrm{w} \in \mathcal{W}^N : \sup_{\theta\in\Theta} |E_g(\theta) - \hat{E}_g(\theta,\mathrm{w})| > \epsilon\right\} = q_g(N,\epsilon). \end{aligned}$$

The second inequality is now proved. Using the definitions of $p_g(\cdot,\cdot,\cdot)$ and $r_g(\cdot,\cdot)$, it suffices to show that

$$\frac{E_g(\theta) - \hat{E}_g(\theta,\mathrm{w})}{\sqrt{E_g(\theta)}} \leq \frac{\epsilon}{\sqrt{\epsilon+\rho}} \quad \text{and} \quad \hat{E}_g(\theta,\mathrm{w}) \leq \rho \qquad (22)$$

implies $E_g(\theta) \leq \hat{E}_g(\theta,\mathrm{w}) + \epsilon$. Letting $\Delta = E_g(\theta) - \hat{E}_g(\theta,\mathrm{w})$ and $\tau = \epsilon/\sqrt{\epsilon+\rho}$, the first inequality of (22) can be rewritten as

$$\frac{\Delta}{\sqrt{\Delta + \hat{E}_g(\theta,\mathrm{w})}} \leq \tau.$$

Clearly, there exists $\mu \geq 0$ such that

$$\frac{\Delta}{\sqrt{\Delta + \hat{E}_g(\theta,\mathrm{w})}} = \tau - \mu.$$

Using this equation, we have

$$\begin{aligned} &\frac{\Delta^2}{\Delta + \hat{E}_g(\theta,\mathrm{w})} = (\tau-\mu)^2 \\ &\Delta^2 - (\tau-\mu)^2\Delta - (\tau-\mu)^2\hat{E}_g(\theta,\mathrm{w}) = 0 \\ &\Delta = \frac{(\tau-\mu)^2 \pm \sqrt{(\tau-\mu)^4 + 4(\tau-\mu)^2\hat{E}_g(\theta,\mathrm{w})}}{2} \\ &\Delta \leq \frac{(\tau-\mu)^2 + \sqrt{(\tau-\mu)^4 + 4(\tau-\mu)^2\hat{E}_g(\theta,\mathrm{w})}}{2}. \end{aligned}$$

Bearing in mind that $\mu \geq 0$ and $\hat{E}_g(\theta,\mathrm{w}) \in [0,\rho]$ we easily obtain the relation

$$\begin{aligned} \Delta \leq \frac{\tau^2 + \sqrt{\tau^4 + 4\tau^2\rho}}{2} &= \frac{\tau^2 + \tau\sqrt{\tau^2+4\rho}}{2} \\ &= \frac{1}{2}\left(\tau^2 + \tau\sqrt{\frac{\epsilon^2}{\epsilon+\rho} + 4\rho}\right) \\ &= \frac{1}{2}\left(\tau^2 + \tau\sqrt{\frac{\epsilon^2 + 4\rho\epsilon + 4\rho^2}{\epsilon+\rho}}\right) \\ &= \frac{1}{2}\left(\tau^2 + \frac{\tau(\epsilon+2\rho)}{\sqrt{\epsilon+\rho}}\right) \\ &= \frac{1}{2}\left(\frac{\epsilon^2}{\epsilon+\rho} + \frac{\epsilon(\epsilon+2\rho)}{\epsilon+\rho}\right) \\ &= \frac{1}{2}\left(\frac{2\epsilon(\epsilon+\rho)}{\epsilon+\rho}\right) = \epsilon. \end{aligned}$$

■

APPENDIX II
PROOF OF THEOREM 6

Clearly, inequality (14) is equivalent to

$$\begin{aligned} \ln a + d\ln \frac{ceN}{d} - bN &< \ln \delta \\ bN - d\ln \frac{ceN}{d} &> \ln \frac{a}{\delta} \\ \frac{bN}{d} - \ln \frac{ceN}{d} &> \frac{1}{d}\ln \frac{a}{\delta}. \end{aligned}$$

Letting $\gamma = bN/d$, it easily follows that the inequality (14) is equivalent to

$$\begin{aligned} \gamma - \ln \frac{ce\gamma}{b} &> \frac{1}{d}\ln \frac{a}{\delta} \\ \gamma - \ln \gamma &> \frac{1}{d}\ln \frac{a}{\delta} + \ln \frac{ce}{b}. \end{aligned} \tag{23}$$

Next, from the convexity of $-\ln \gamma$, we notice that $-\ln \gamma$ is no smaller than its linearization at $\gamma = \mu$. Therefore, the inequality (23) holds if there exists $\mu > 1$ such that

$$\begin{aligned} \gamma - \ln \mu - \frac{1}{\mu}(\gamma - \mu) &> \frac{1}{d}\ln \frac{a}{\delta} + \ln \frac{ce}{b} \\ \left(1 - \frac{1}{\mu}\right)\gamma - \ln \mu + 1 &> \frac{1}{d}\ln \frac{a}{\delta} + \ln \frac{ce}{b} \\ \left(1 - \frac{1}{\mu}\right)\gamma &> \frac{1}{d}\ln \frac{a}{\delta} + \ln \frac{c\mu}{b} \\ \gamma &> \left(\frac{\mu}{\mu - 1}\right)\left(\frac{1}{d}\ln \frac{a}{\delta} + \ln \frac{c\mu}{b}\right). \end{aligned}$$

Since $\gamma = bN/d$ we conclude that the inequality (14) is satisfied if

$$N \geq \inf_{\mu > 1} \frac{\mu}{b(\mu - 1)}\left(\ln \frac{a}{\delta} + d\ln \frac{c\mu}{b}\right). \tag{24}$$

Recall now that $a \geq 1$, $b \in (0,1]$, $c \geq 1$, $d \geq 1$ and $\delta \in (0,1)$. Then, we conclude that inequality (24) implies

$$N \geq \inf_{\mu > 1}\left(\frac{\mu}{(\mu - 1)}\ln \mu\right)d.$$

Finally, note that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{\mu}{(\mu - 1)}\ln \mu\right) &= \left(\frac{\mu - 1 - \mu}{(\mu - 1)^2}\right)\ln \mu + \frac{1}{\mu - 1} \\ &= \left(\frac{-1}{(\mu - 1)^2}\right)\ln(1 + (\mu - 1)) \\ &\quad + \frac{1}{\mu - 1} \\ &\geq \frac{1 - \mu}{(\mu - 1)^2} + \frac{1}{\mu - 1} = 0. \end{aligned}$$

Using this fact, we conclude that $d\mu/(\mu - 1)\ln \mu$ is a strictly growing function for $\mu > 1$. This means that

$$N \geq \inf_{\mu > 1}\left(\frac{\mu}{(\mu - 1)}\ln \mu\right)d = \lim_{\mu \to 1}\left(\frac{\mu}{(\mu - 1)}\ln \mu\right)d = d.$$

■

APPENDIX III
A PACK-BASED STRATEGY AND PROOF OF THEOREM 8

In this appendix, a technique introduced recently (see [1], [3]), denoted as *pack-based strategy*, is presented. This allows us to obtain improved sample size bounds for the particular case when the level parameter ρ is equal to zero. This technique is also instrumental to the proof of Theorem 8.

Given an integer L, we say that z belongs to the sample space $\mathcal{S}_L$ if z belongs to the Cartesian product $\mathcal{W} \times \cdots \times \mathcal{W}$ (L times). Then, given $g : \Theta \times \mathcal{W} \to \{0,1\}$, the function $g_L : \Theta \times \mathcal{S}_L \to \{0,1\}$ is defined as follows. For a given $\theta \in \Theta$ and a collection of L samples $z = \{w^{(1)}, \ldots, w^{(L)}\} \in \mathcal{S}_L$,

$$g_L(\theta, z) = \max_{i=1,\ldots,L} g(\theta, w^{(i)}). \tag{25}$$

The following notation is introduced to emphasize that given a collection of $N = LM$ samples, the multisample z may be considered as a collection of M packs of L samples each. Given positive integers L and M, the collection $\mathrm{z} = \{z^{(1)}, \ldots, z^{(M)}\}$ is said to belong to the set $\mathcal{S}_L^M$ if $z^{(i)} \in \mathcal{S}_L$, $i = 1, \ldots, M$. Note that there is an equivalence one-to-one between the elements of $\mathcal{W}^{LM} = \mathcal{W}^N$ and the elements of $\mathcal{S}_L^M$.

The notion of probability of violation introduced in Definition 1, is now generalized to the function $g_L(\cdot,\cdot)$

$$E_{g_L}(\theta) = \mathrm{Pr}_{\mathcal{S}_L}\{z \in \mathcal{S}_L : g_L(\theta, z) = 1\}.$$

Given $\mathrm{z} = \{z^{(1)}, \ldots, z^{(M)}\} \in \mathcal{S}_L^M$, and $\theta \in \Theta$, the empirical mean of $g_L(\theta, z)$ with respect to the multisample z is defined as

$$\hat{E}_{g_L}(\theta, \mathrm{z}) := \frac{1}{M}\sum_{i=1}^{M} g_L(\theta, z^{(i)}). \tag{26}$$

The following lemma, which is proved in Appendix 4, shows the relationships between the probabilities $E_{g_L}(\theta)$ and $E_g(\theta)$. We also remark that the "log-over-log" bound appearing here has been stated in [31], see also [32].

Lemma 3: Suppose that $\theta \in \Theta$ satisfies $E_{g_L}(\theta) \leq \tau < 1$. Assume that $\epsilon \in (0,1)$ is given and

$$L \geq \frac{\ln(1 - \tau)}{\ln(1 - \epsilon)}.$$

Then, we have that $E_g(\theta) \leq \epsilon$.

We are now ready to introduce the following *pack-based zero-level randomized strategy*:

(i) Given positive integers L and M, draw $N = LM$ independent identically distributed samples $\{w^{(1)}, \ldots, w^{(N)}\}$ according to the probability $\mathrm{Pr}_{\mathcal{W}}$.

(ii) Pack the N samples as follows:

$$\begin{aligned} z^{(1)} &= (w^{(1)}, w^{(2)}, \ldots, w^{(L)}) \\ z^{(2)} &= (w^{(1+L)}, w^{(2+L)}, \ldots, w^{(2L)}) \\ \vdots &= \vdots \\ z^{(i)} &= (w^{(1+(i-1)L)}, w^{(2+(i-1)L)}, \ldots, w^{(iL)}) \\ \vdots &= \vdots \\ z^{(M)} &= (w^{(1+(M-1)L)}, w^{(2+(M-1)L)}, \ldots, w^{(ML)}). \end{aligned}$$

(iii) Find (if possible) a feasible solution $\theta \in \Theta$ to the pack-based feasibility problem

$$\hat{E}_{g_L}(\theta, \mathrm{z}) = 0. \tag{27}$$

(iv) If this problem is feasible, find a solution to the pack-based optimization problem

$$\min_{\theta\in\Theta} J(\theta) \text{ subject to } \hat{E}_{g_L}(\theta, \mathrm{z}) = 0. \tag{28}$$

We introduce a simple result which states the equivalence between problems (16) and (28).

Lemma 4: Given L and M, suppose that $N = LM$. Then, given the multi-sample extraction $\{w^{(1)}, \ldots, w^{(N)}\}$, the problem (16) is feasible if and only if the problem (28) is feasible. In case of feasibility, $\hat{\theta}_N = \hat{\theta}_{L,M}$ where $\hat{\theta}_N$ and $\hat{\theta}_{L,M}$ denote the optimal solutions of problems (16) and (28), respectively.

Proof: To prove the result it suffices to notice that problems (16) and (28) yield the same optimization problem. ■

We are now ready to prove Theorem 8. To this end, note that the definition of probability of two-sided failure applies to the function $g_L(\cdot,\cdot)$ in a natural way. Given M, $\tau \in (0,1)$ and the function $g_L : \Theta \times \mathcal{S}_L \to \{0,1\}$, the probability of two-sided failure $q_{g_L}(M,\tau)$ is defined by

$$q_{g_L}(M,\tau) := \mathrm{Pr}_{\mathcal{S}_L^M}\left\{\mathrm{z} \in \mathcal{S}_L^M : \sup_{\theta\in\Theta} |E_{g_L}(\theta) - \hat{E}_{g_L}(\theta,\mathrm{z})| > \tau\right\}.$$

An intermediate result is now stated.

Lemma 5: Given $\epsilon \in (0,1)$ and $\delta \in (0,1)$, suppose that M, L and N are chosen such that $q_{g_L}(M,\tau) \le \delta$, $L \ge \ln(1-\tau)/\ln(1-\epsilon)$ and $N = LM$. Then, with probability no smaller than $1-\delta$, either the zero-level randomized optimization problem (16) is infeasible and, hence, also the general optimization problem (2) is infeasible; or, (16) is feasible, and then any feasible solution θ satisfies the inequality $E_g(\theta) \le \epsilon$.

Proof: If the zero-level randomized optimization problem (16) is infeasible, then the general optimization problem (2) is also infeasible. Consider now the case where problem (16) is feasible.

Clearly, if θ is a feasible solution to problem (16), then it is also a feasible solution to the optimization problem (28). This implies that $\hat{E}_{g_L}(\theta,\mathrm{z}) = 0$. Denote by γ the probability that $E_g(\theta) > \epsilon$. Lemma 3 allows one to affirm that this probability γ is no larger than the probability of $E_{g_L}(\theta) > \tau$. We now prove that γ is no larger than δ

$$\begin{aligned}\gamma &\le \mathrm{Pr}_{\mathcal{S}_L^M}\left\{\mathrm{z} \in \mathcal{S}_L^M : \sup_{\hat{E}_{g_L}(\theta,\mathrm{z})=0, \theta\in\Theta} E_{g_L}(\theta) > \tau\right\}\\ &\le \mathrm{Pr}_{\mathcal{S}_L^M}\left\{\mathrm{z} \in \mathcal{S}_L^M : \sup_{\theta\in\Theta} |E_{g_L}(\theta) - \hat{E}_{g_L}(\theta,\mathrm{z})| > \tau\right\}\\ &= q_{g_L}(M,\tau) \le \delta.\end{aligned}$$

■

We now continue with the proof of Theorem 8. Suppose that $\mathrm{VC}_{g_L} \le d_L$ and that $\tau \in (0,1)$. As $\mathrm{VC}_{g_L} \le d_L$, we infer from Corollary 2 that if $M > d_L$, then

$$q_{g_L}(M,\tau) \le 4e^{2\tau}\left(\frac{2eM}{d_L}\right)^{d_L} e^{-M\tau^2}.$$

Invoking Theorem 6 for $a = 4e^{2\tau}$, $b = \tau^2$ and $c = 2$ we obtain that $q_{g_L}(M,\tau) \le \delta$ if

$$M \ge \inf_{\mu>1} \frac{\mu}{\tau^2(\mu-1)}\left(\ln\frac{4e^{2\tau}}{\delta} + d_L \ln\frac{2\mu}{\tau^2}\right).$$

This means, by Lemma 5, that the claim of Theorem 8 holds if $N = ML$ where

$$\begin{aligned}M &\ge \inf_{\mu>1} \frac{\mu}{\tau^2(\mu-1)}\left(\ln\frac{4e^{2\tau}}{\delta} + d_L \ln\frac{2\mu}{\tau^2}\right),\\ L &\ge \left\lceil\frac{\ln(1-\tau)}{\ln(1-\epsilon)}\right\rceil.\end{aligned}$$

Equivalently, the number of samples N required to guarantee that the claim of Theorem 8 holds is bounded by

$$N \ge \inf_{\mu>1} \left\lceil\frac{\ln(1-\tau)}{\ln(1-\epsilon)}\right\rceil \frac{\mu}{\tau^2(\mu-1)}\left(\ln\frac{4e^{2\tau}}{\delta} + d_L \ln\frac{2\mu}{\tau^2}\right). \tag{29}$$

From the assumptions of the theorem it is easy to see that $g_L(\cdot,\cdot)$ can be written as an (α, Lk)-Boolean function (see Definition 7). Therefore, using Lemma 2 we obtain that $\mathrm{VC}_{g_L} \le 2n_\theta \log_2(4e\alpha Lk)$. Thus, $d_L = 2n_\theta \log_2(4e\alpha Lk)$ can be taken as an upper bound of the VC-dimension of $g_L(\cdot,\cdot)$.

To minimize the number of samples we fix τ and μ to 0.844 and 3.193, respectively. These values correspond to the numerical minimization of the term $\left(\ln(1-\tau)\mu/\tau^2(\mu-1)\right) \ln 2\mu/\tau^2$. This choice of μ and τ and the inequality $-\ln(1-\epsilon) > \epsilon$ lead to the following bound:

$$N \ge 2.05\left\lceil\frac{1.86}{\epsilon}\right\rceil\left(\ln\frac{21.64}{\delta} + 4.39 n_\theta \log_2\left(4e\alpha k\left\lceil\frac{1.86}{\epsilon}\right\rceil\right)\right).$$

Since $\epsilon \in (0, 0.14)$, we have

$$\left\lceil\frac{1.86}{\epsilon}\right\rceil \le \frac{1.86+\epsilon}{\epsilon} \le \frac{1.86+0.14}{\epsilon} = \frac{2}{\epsilon}. \tag{30}$$

Substituting this inequality in the previous bound we obtain

$$N \ge \frac{4.1}{\epsilon}\left(\ln\frac{21.64}{\delta} + 4.39 n_\theta \log_2\left(\frac{8e\alpha k}{\epsilon}\right)\right).$$

■

Appendix IV
Proof of Lemma 3

Two cases should be considered, i.e., $\tau \le \epsilon$ and $\tau > \epsilon$. The proof for the first case is trivial because $E_g(\theta) \le E_{g_L}(\theta)$, for all $L \ge 1$. This implies that $E_g(\theta) \le E_{g_L}(\theta) \le \tau \le \epsilon$. In what follows, the second case will be analyzed. We recall that the probability of violation is given by

$$E_g(\theta) = \mathrm{Pr}_{\mathcal{W}}\{w \in \mathcal{W} : g(\theta, w) = 1\}.$$

This means that the probability of drawing a collection of L samples $\mathrm{w} = \{w^{(1)}, \ldots, w^{(L)}\}$ and obtaining

$$g_L(\theta, \mathrm{w}) = \max_{i=1,\ldots,L} g(\theta, w^{(i)}) = 0$$

is equal to $(1 - E_g(\theta))^L$. From this fact, we conclude that $E_{g_L}(\theta) = 1 - (1 - E_g(\theta))^L$. As $E_{g_L}(\theta) \leq \tau$ it follows that

$$1 - \tau \leq (1 - E_g(\theta))^L. \tag{31}$$

Since $(1 - E_g(\theta)) \in [0, 1]$ and L satisfies the inequality

$$L \geq \frac{\ln(1-\tau)}{\ln(1-\epsilon)},$$

we conclude from (31) that

$$1 - \tau \leq (1 - E_g(\theta))^{\ln(1-\tau)/\ln(1-\epsilon)}.$$

Applying the natural logarithm to both sides of the inequality, we have

$$\ln(1-\tau) \leq \left(\frac{\ln(1-\tau)}{\ln(1-\epsilon)}\right) \ln(1 - E_g(\theta)).$$

Multiplying both sides of this inequality by $\ln(1-\epsilon)/\ln(1-\tau) > 0$, the following sequence of relations is obtained

$$\begin{aligned} \ln(1-\epsilon) &\leq \ln(1 - E_g(\theta)) \\ 1-\epsilon &\leq 1 - E_g(\theta) \\ E_g(\theta) &\leq \epsilon. \end{aligned}$$

■

APPENDIX V
PROOF OF THE STATEMENT IN PROPERTY 1

First, let $\mathrm{v}_k := \{v^{(1)}, \ldots, v^{(M_k)}\}$. Denote also by δ_k the probability of obtaining $\hat{\theta}_{N_k}$ satisfying $\hat{E}_g(\hat{\theta}_{N_k}, \mathrm{v}_k) \leq \rho + (1 - \beta_v^{-k/2})\epsilon$ and $E_g(\hat{\theta}_{N_k}) > \rho + \epsilon$ at iteration k, $k \leq k_t - 1$. Clearly, this probability satisfies

$$\delta_k \leq \mathrm{Pr}_{\mathcal{W}^{M_k}}\{\mathrm{w} \in \mathcal{W}^{M_k} : \hat{E}_g(\hat{\theta}_{N_k}, \mathrm{w}) \leq \rho + (1 - \beta_v^{-k/2})\epsilon \quad \text{and} \quad E_g(\hat{\theta}_{N_k}) > \rho + \epsilon\}.$$

Next define $\rho_k := \rho + (1 - \beta_v^{-k/2})\epsilon$ and $\epsilon_k := \beta_v^{-k/2}\epsilon$. With this notation, it results that $\epsilon_k \in (0, 1)$ and consequently $\epsilon_k/\sqrt{\epsilon_k + \rho_k} \in (0, 1)$. Therefore, the previous bound on δ_k is rewritten as

$$\delta_k \leq \mathrm{Pr}_{\mathcal{W}^{M_k}}\{\mathrm{w} \in \mathcal{W}^{M_k} : \hat{E}_g(\hat{\theta}_{N_k}, \mathrm{w}) \leq \rho_k \quad \text{and} \quad E_g(\hat{\theta}_{N_k}) > \hat{E}_g(\hat{\theta}_{N_k}, \mathrm{w}) + \epsilon_k\}.$$

Taking into account the statement presented in (22) in the proof of Theorem 1, it follows that:

$$\delta_k \leq \mathrm{Pr}_{\mathcal{W}^{M_k}}\left\{\mathrm{w} \in \mathcal{W}^{M_k} : \frac{E_g(\hat{\theta}_{N_k}) - \hat{E}_g(\hat{\theta}_{N_k}, \mathrm{w})}{\sqrt{E_g(\hat{\theta}_{N_k})}} > \frac{\epsilon_k}{\sqrt{\epsilon_k + \rho_k}}\right\}. \tag{32}$$

The multiplicative Chernoff bound [34] states that given a fixed θ and $\gamma \in (0, 1)$,

$$\mathrm{Pr}_{\mathcal{W}^N}\left\{\mathrm{w} \in \mathcal{W}^N : \frac{E_g(\theta) - \hat{E}_g(\theta, \mathrm{w})}{\sqrt{E_g(\theta)}} > \gamma\right\} < e^{-\gamma^2 N/2}.$$

Combining this inequality with (32) we obtain

$$\ln \delta_k \leq \frac{-\epsilon_k^2 M_k}{2(\epsilon_k + \rho_k)} = \frac{-\beta_v^{-k}\epsilon^2 M_k}{2(\epsilon + \rho)}.$$

Therefore, the proposed choice for M_k guarantees that

$$\ln \delta_k \leq \frac{-\beta_v^{-k}\epsilon^2}{2(\epsilon + \rho)} \left\lceil 2\beta_v^k \left(\frac{\rho + \epsilon}{\epsilon^2}\right) \ln \frac{2k_t}{\delta} \right\rceil \leq \ln \frac{\delta}{2k_t}.$$

Finally, we conclude that the probability of misclassification δ_k is smaller than $\delta/2k_t$, for $k = 1, \ldots, k_t - 1$. Let δ_{k_t} denote the probability of misclassification under the assumption that k reaches the value k_t. This probability is determined by the choice of $\hat{N}_t$,

$$\hat{N}_t = \frac{5(\rho + \epsilon)}{\epsilon^2}\left(\ln \frac{8}{\delta} + d \ln \frac{40(\rho + \epsilon)}{\epsilon^2}\right).$$

We easily obtain from Theorem 7 that $p_g(\hat{N}_t, \epsilon, \rho) \leq \delta/2$. Since the selection of β_p guarantees that $N_{k_t} = \hat{N}_t$, the probability of misclassification at iteration k_t is no larger than $\delta/2$. We finally conclude that the probability of misclassification of the algorithm is no larger than

$$\sum_{k=1}^{k_t - 1} \delta_k + \delta_{k_t} \leq \frac{\delta(k_t - 1)}{2k_t} + \frac{\delta}{2} \leq \delta.$$

We notice that this bound on the probability of misclassification is related to that presented in [25], which relies on the convergence of the series $\sum_{k=1}^{\infty} 1/k^2$. ■

APPENDIX VI
PROOF OF PROPERTY 2

To prove this property we proceed by contradiction. Suppose that there exists $\hat{\theta}$ such that $E_g(\hat{\theta}) \leq \rho - \mu$ and that the algorithm reaches iteration k_t. In addition, suppose that problem (20) is infeasible also for this last iteration with probability greater than δ. Then, we bound the probability τ that at iteration k_t the problem (20) is infeasible for the i.i.d. multisample $\mathrm{w}_{k_t} = \{w^{(1)}, \ldots, w^{(\hat{N}_t)}\}$. Since we assume that there exists at least an element $\hat{\theta} \in \Theta$ satisfying $E_g(\hat{\theta}) \leq \rho - \mu$ and $N_{k_t} = \hat{N}_t$, this probability is no larger than $\tau = \mathrm{Pr}_{\mathcal{W}^{\hat{N}_t}}\{\mathrm{w} \in \mathcal{W}_{\hat{N}_t} :$

$\hat{E}_g(\hat{\theta},\mathrm{w}) > E_g(\hat{\theta})+\mu\}$. Finally, using the Hoeffding inequality provided in (8) we obtain

$$\tau \leq 2e^{-2\hat{N}_t\mu^2} = 2e^{-2\hat{N}_t(\frac{1}{2\hat{N}_t}\ln\frac{2}{\delta})} = \delta.$$

■

References

[1] T. Alamo, R. Tempo, and E. F. Camacho, "Improved sample size bounds for probabilistic robust control design: A pack-based strategy," in *Proc. 46th IEEE Conf. Decision Control*, New Orleans, LA, Dec. 2007, pp. 6178–6183.

[2] T. Alamo, R. Tempo, and E. F. Camacho, "Revisiting statistical learning theory for uncertain feasibility and optimization problems," in *Proc. 46th IEEE Conf. Decision Control*, New Orleans, LA, Dec. 2007, pp. 2135–2140.

[3] T. Alamo, R. Tempo, and E. F. Camacho, "Statistical learning theory: A pack-based strategy for uncertain feasibility and optimization problems," in *Recent Advances in Learning and Control*, V. D. Blondel, S. P. Boyd, and H. Kimura, Eds. London, U.K.: Springer Verlag, 2008.

[4] T. Alamo, R. Tempo, D. R. Ramirez, and E. F. Camacho, "A new vertex result for robustness problems with interval matrix uncertainty," *Syst. Control Lett.*, vol. 57, pp. 474–481, 2008.

[5] T. Alamo, R. Tempo, D. Rodríguez, and E. F. Camacho, "A sequentially optimal randomized algorithm for robust LMI feasibility problems," in *Proc. Eur. Control Conf.*, Kos, Greece, Jul. 2007, pp. 3620–3627.

[6] B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[7] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA: SIAM, 1999.

[8] E.-W. Bai, R. Tempo, and M. Fu, "Worst-case properties of the uniform distribution and randomized algorithms for robustness analysis," *Math. Control, Signals, Syst.*, vol. 11, pp. 183–196, 1998.

[9] B. R. Barmish and C. M. Lagoa, "The uniform distribution: A rigorous justification for its use in robustness analysis," *Math. Control, Signals, Syst.*, vol. 10, pp. 203–222, 1997.

[10] V. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in system and control," *Automatica*, vol. 36, pp. 1249–1274, 2000.

[11] G. Calafiore and M. C. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Math. Program.*, vol. 102, pp. 25–46, 2005.

[12] G. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, May 2006.

[13] G. Calafiore and F. Dabbene, "A probabilistic analytic center cutting plane method for feasibility of uncertain LMIs," *Automatica*, vol. 43, pp. 2022–2033, 2007.

[14] G. Calafiore and B. T. Polyak, "Stochastic algorithms for exact and approximate feasibility of robust LMIs," *IEEE Trans. Autom. Control*, vol. 46, no. 11, pp. 1755–1759, Nov. 2001.

[15] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of robust convex programs," *SIAM J. Optim.*, vol. 19, pp. 1211–1230, 2008.

[16] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals Math. Stat.*, vol. 23, pp. 493–507, 1952.

[17] Y. Fujisaki, F. Dabbene, and R. Tempo, "Probabilistic design of LPV control systems," *Automatica*, vol. 39, pp. 1323–1337, 2003.

[18] Y. Fujisaki and Y. Kozawa, "Probabilistic robust controller design: Probable near minimax value and randomized algorithms," in *Probabilistic and Randomized Methods for Design under Uncertainty*, G. Calafiore and F. Dabbene, Eds. London, U.K.: Springer Verlag, 2006.

[19] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, pp. 13–30, 1963.

[20] H. Ishii, T. Basar, and R. Tempo, "Randomized algorithms for synthesis of switching rules for multimodal systems," *IEEE Trans. Autom. Control*, vol. 50, no. 6, pp. 754–767, Jun. 2005.

[21] S. Kanev, B. De Schutter, and M. Verhaegen, "An ellipsoid algorithm for probabilistic robust controller design," *Syst. Control Lett.*, vol. 49, pp. 365–375, 2003.

[22] M. Karpinski and A. Macintyre, "Polynomial bounds for VC dimension of sigmoidal and general pfaffian neural networks," *J. Comput. Syst. Sci.*, vol. 54, pp. 169–176, 1997.

[23] V. Koltchinskii, C. T. Abdallah, M. Ariola, P. Dorato, and D. Panchenko, "Improved sample complexity estimates for statistical learning control of uncertain systems," *IEEE Trans. Autom. Control*, vol. 45, no. 12, pp. 2383–2388, Dec. 2000.

[24] S. Mendelson, "A few notes on statistical learning theory," in *Advanced Lectures in Machine Learning*, S. Mendelson and A. J. Smola, Eds. London, U.K.: Springer Verlag, 2003.

[25] Y. Oishi, "Polynomial-time algorithms for probabilitic solutions of parameter-dependent linear matrix inequalities," *Automatica*, vol. 43, pp. 538–545, 2007.

[26] J. M. Parrondo and C. van den Broeck, "Vapnik-Chervonenkis bounds for generalization," *J. Phys. A*, vol. 26, pp. 2211–2223, 1993.

[27] B. T. Polyak and R. Tempo, "Probabilistic robust design with linear quadratic regulators," *Syst. Control Lett.*, vol. 43, pp. 343–353, 2001.

[28] N. Sauer, "On the density of family of sets," *J. Combinatorial Theory*, vol. 13, no. A, pp. 145–147, 1972.

[29] C. Scherer, "LMI relaxations in robust control," *Eur. J. Control*, vol. 12, no. 1, pp. 3–29, 2006.

[30] E. D. Sontag, "VC dimension of neural networks," in *Neural Networks and Machine Learning*, C. M. Bishop, Ed. New York: Springer Verlag, 1998.

[31] R. Tempo, E.-W. Bai, and F. Dabbene, "Probabilistic robustness analysis: Explicit bounds for the minimum number of samples," *Syst. Control Lett.*, vol. 30, pp. 237–242, 1997.

[32] R. Tempo, G. Calafiore, and F. Dabbene, "Randomized algorithms for analysis and control of uncertain systems," in *Communications and Control Engineering Series*. London, U.K.: Springer Verlag, 2005.

[33] J. S. Tyler and F. B. Tuteur, "The use of a quadratic performance index to design multivariable invariant plants," *IEEE Trans. Autom. Control*, vol. 11, no. 1, pp. 84–92, Jan. 1966.

[34] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[35] V. N. Vapnik and A. Ya. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory Probab. Appl.*, vol. 16, pp. 264–280, 1971.

[36] V. N. Vapnik and A. Ya. Chervonenkis, "Necessary and sufficient conditions for the uniform convergence of means to their expectations," *Theory Probab. Appl.*, vol. 26, no. 3, pp. 532–553, 1981.

[37] M. Vidyasagar, *A Theory of Learning and Generalization: With Applications to Neural Networks and Control Systems*. London, U.K.: Springer Verlag, 1997.

[38] M. Vidyasagar, "Statistical learning theory and randomized algorithms for control," *Control Syst. Mag.*, vol. 18, no. 6, pp. 69–85, 1998.

[39] M. Vidyasagar, "Randomized algorithms for robust controller synthesis using statistical learning theory," *Automatica*, vol. 37, pp. 1515–1528, 2001.

[40] M. Vidyasagar and V. D. Blondel, "Probabilistic solutions to some NP-hard matrix problems," *Automatica*, vol. 37, no. 9, pp. 1397–1405, 2001.

Teodoro Alamo (M'98) was born in Spain in 1968. He received the M.Eng. degree in telecommunications engineering from the Polytechnic University of Madrid, Madrid, Spain, in 1993 and the Ph.D. degree in telecommunications engineering from the University of Seville, Seville, Spain, in 1998.

From 1993 to 2000, he was an Assistant Professor of the Department of System Engineering and Automatic Control, University of Seville. Since 2001, he has been an Associate Professor in the same department. He was a Researcher in the Ecole Nationale Supérieure des Télécommunications (Telecom Paris) from 1991 to 1993 and he has participated in several European Projects. He is the author or coauthor of more than 120 publications including book chapters, journal papers, conference proceedings and educational books. He has carried out reviews for various conferences and technical journals. His current research interests are in model predictive control, randomized algorithms, robust control, identification, control of constrained systems, invariant sets and convex optimization.

Roberto Tempo (F'00) is currently a Director of Research of Systems and Computer Engineering at the institute IEIIT, National Research Council (CNR), Politecnico di Torino, Italy. He has held visiting positions at various institutions, including Kyoto University, University of Illinois at Urbana-Champaign, German Aerospace Research Organization in Oberpfaffenhofen and Columbia University in New York. He is author or co-author of more than 150 research papers published in international journals, books and conferences. He is also a co-author of *Randomized Algorithms for Analysis and Control of Uncertain Systems* (New York: Springer-Verlag, 2005). His research activities are mainly focused on analysis and design of complex systems with uncertainty, and related control applications.

Dr. Tempo received the "Outstanding Paper Prize Award" from the International Federation of Automatic Control (IFAC) for a paper published in *Automatica*, and of the "Distinguished Member Award" from the IEEE Control Systems Society. He is a Fellow of the IEEE and a Fellow of the IFAC. He is currently an Editor and Deputy Editor-in-Chief of *Automatica*, and he has been Editor for Technical Notes and Correspondence of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He has served as Program Chair of the first joint IEEE Conference on Decision and Control and European Control Conference, Seville, Spain, 2005. He is the 2009 President-Elect of the IEEE Control Systems Society.

Eduardo F. Camacho (M'85–SM'04) received the Ph.D. degree from the University of Seville, Seville, Spain.

He is a Professor in the Department of System Engineering and Automatic Control, University of Seville. He has co-authored *Model Predictive Control in the Process industry* (1995), *Advanced Control of Solar Plants* (1997), *Model Predictive Control* (1999) (New York: Springer–Verlag, 2004, second edition), *Control e Instrumentación de Procesos Químicos* (Ed. Síntesis), and *Control of Dead time Processes* (New York: Springer-Verlag, 2007). He has authored and co-authored more than 200 technical papers in international journals and conference proceedings.

Dr. Camacho has served on various IFAC and IEEE/CSS committees. He chaired the IEEE/CSS International Affairs Committee and the IFAC Publication Committee and has been the president of the European Union Control Association (EUCA). At present he chairs the IFAC Policy Committee and is a member of the IEEE/CSS Board of Governors and vice president of EUCA. He has carried out review and editorial work for various conferences and technical journals. He is one of the editors of the IFAC journal, Control Engineering Practice, of the European Journal of Control and of the Optimal Control: Application and Methods. He was Publication Chair for the IFAC World Congress B'02 and he has been the General Chair of the joint 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005. He has acted as an evaluator of projects at national and European level. He was appointed for four years as the Manager of the Advanced Production Technology Program of the Spanish National R&D Program. He was one of the Spanish representatives on the Program Committee of the Growth Research program and for the Program Committee of the NMP research priority of the European Union.