43rd IEEE Conference on Decision and Control
December 14-17, 2004
Atlantis, Paradise Island, Bahamas

A new bound on the generalization rate of sampled convex programs

Giuseppe Calafiore and M.C. Campi

***Abstract*— This paper deals with the sampled scenarios approach to robust convex programming. It has been shown in previous works that by randomly sampling a sufficient number of constraints among the (possibly) infinite constraints of a robust convex program, one obtains a standard convex optimization problem whose solution is 'approximately feasible', in a probabilistic sense, for the original robust convex program. This is a *generalization* property in the learning theoretic sense, since the satisfaction of a certain number of 'training' constraints entails the satisfaction of other 'unseen' constraints. In this paper we provide a new efficient bound on the generalization rate of sampled convex programs, and show an example of application to a robust control design problem.**

***Keywords*: Uncertain convex optimization, Robust control, Randomized algorithms, Probabilistic robustness.**

I. Introduction

Robust convex programming [3], [18] deals with optimization problems subject to a family of convex constraints that are parameterized by uncertainty terms. Solving a robust convex program (RCP) amounts to determining an optimal solution that is feasible for all possible constraints in the parameterized family. In more precise terms, a RCP may be formalized as

$$\mathrm{RCP}: \quad \min_{\theta\in\Theta} c^T\theta \quad \text{subject to:} \quad f(\theta,\delta)\leq 0, \quad \forall\delta\in\Delta, \tag{1}$$

where θ is the optimization variable, δ is the uncertainty parameter, $\Theta\subseteq\mathbb{R}^{n_\theta}$ is a convex and closed set, and $\Delta\subseteq\mathbb{R}^{n_\delta}$. Further, it is assumed that $f(\theta,\delta):\Theta\times\Delta\to(-\infty,\infty]$ is continuous and convex in θ, for any fixed value of $\delta\in\Delta$. Notice that no assumption is instead made on the dependence of $f(\theta,\delta)$ on δ.

The constraints are here expressed by the condition $f(\theta,\delta)\leq 0$, where f is a scalar-valued function. Considering scalar-valued constraint functions is without loss of generality, since multiple constraints $f_1(\theta,\delta)\leq 0,\ldots,f_{n_f}(\theta,\delta)\leq 0$ can be reduced to a single scalar-valued constraint by the position $f(\theta,\delta)\doteq\max_{i=1,\ldots,n_f} f_i(\theta,\delta)$.

We remark that, despite convexity, robust convex programs are in general NP-hard, see [3], [5], [18]. This is one of the motivations that led us to consider probabilistic relaxations of the problem, see [7] for an in-depth discussion.

Important special cases of robust convex programs are robust linear programs, [4], for which $f(\theta,\delta)=\max_{i=1,\ldots,n_f} f_i(\theta,\delta)$, where each $f_i(\theta,\delta)$ is affine in θ, and robust semidefinite programs, [1], [5], [18], for which $f(\theta,\delta)=\lambda_{\max}[F(\theta,\delta)]$, where

$$F(\theta,\delta)=F_0(\delta)+\sum_{i=1}^{n_\theta}\theta_i F_i(\delta), \quad F_i(\delta)=F_i^T(\delta),$$

and $\lambda_{\max}[\cdot]$ denotes the largest eigenvalue.

The RCP paradigm has found to date applications in many engineering endeavours, such as truss topology design [2], robust antenna array design, portfolio optimization [19], and robust estimation [17]. However, we shall here be mainly concerned with control systems, where RCPs arise naturally in the context of analysis and synthesis based on parameter-dependent Lyapunov functions, see e.g. [1], [10], [11], [12], as well as in various problems of robust filtering [15], [25] and set-membership state reachability and filtering [8], [16].

In [6], [7], a probabilistic approach has been proposed to approximately solve problem (1). This approach is based on sampling at random a finite number N of constraints in the family $\{f(\theta,\delta)\leq 0, \delta\in\Delta\}$ and solving the corresponding standard convex problem. In particular, we explicitly define the *scenario* counterpart of RCP as

$$\mathrm{RCP}_N: \quad \min_{\theta\in\Theta} c^T\theta \quad \text{subject to:} \quad f(\theta,\delta^{(i)})\leq 0,\ i=1,\ldots,N, \tag{2}$$

where $\delta^{(1)},\ \ldots,\ \delta^{(N)}$ are N independent identically distributed (iid) samples, drawn according to some given probability measure denoted as 'Prob'. A scenario design is given by an optimal solution $\hat{\theta}_N$ of RCP_N. Notice that $\hat{\theta}_N$ is a random variable that depends on the random extractions $\delta^{(1)},\ldots,\delta^{(N)}$.

A. *Properties of* RCP_N

Let us first specify more precisely our probabilistic setup. We assume that the support Δ for δ is endowed with a σ-algebra $\mathscr{D}$ and that Prob is defined over $\mathscr{D}$. Moreover, we assume that $\{\delta\in\Delta: f(\theta,\delta)\leq 0\}\in\mathscr{D}$, $\forall\theta\in\Theta$. We have the following definition.

Definition 1 (violation probability): Let $\theta\in\Theta$ be given. The *probability of violation* of θ is defined as

$$V(\theta)\doteq\mathrm{Prob}\{\delta\in\Delta:\ f(\theta,\delta)>0\}. \qquad \star$$

This work was supported by FIRB funds of Italian Ministry of University and Research (MIUR), HYBRIDGE IST-2001-32460, and MIUR "New Methods for Identification and Adaptive Control of Industrial Systems".

G. Calafiore is with Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino – Italy. giuseppe.calafiore@polito.it

M.C. Campi is with Dipartimento di Elettronica per l'Automazione, Università di Brescia, Via Branze 38, 25123 Brescia – Italy. campi@ing.unibs.it

0-7803-8682-5/04/$20.00 ©2004 IEEE

For example, if a uniform (with respect to Lebesgue measure) probability distribution is assumed, then $V(\theta)$ measures the volume of 'bad' parameters δ such that the constraint $f(\theta,\delta)\leq 0$ is violated. Clearly, a solution θ with small associated $V(\theta)$ is feasible for most of the problem instances, i.e. it is *approximately feasible* for the robust problem.

Definition 2 (ε-level solution): Let $\varepsilon\in(0,1)$. We say that $\theta\in\Theta$ is an ε-level robustly feasible (or, more simply, an ε-level) solution, if $V(\theta)\leq\varepsilon$. ⋆

Our goal is to devise an algorithm that returns a ε-level solution, where ε is any fixed small level. It was shown in [7] that the solution returned by RCP_N has indeed this characteristic, as summarized in the following theorem.

Theorem 1 (Corollary 1 of [7]): Assume that, for any extraction of $\delta^{(1)}, \ldots, \delta^{(N)}$, the scenario problem RCP_N attains a unique optimal solution $\hat{\theta}_N$.

Fix two real numbers $\varepsilon\in(0,1)$ (level parameter) and $\beta\in(0,1)$ (confidence parameter) and let

$$N\geq N_{\mathrm{lin}}(\varepsilon,\beta)\doteq\left\lfloor\frac{n_\theta}{\varepsilon\beta}\right\rfloor \tag{3}$$

($\lfloor\rfloor$ denotes integer rounding towards zero). Then, with probability no smaller than $1-\beta$, $\hat{\theta}_N$ is ε-level robustly feasible. ⋆

The inequality (3) provides the minimum number of sampled constraints that are needed in order to attain the desired probabilistic levels of robustness in the solution. The function $N_{\mathrm{lin}}(\varepsilon,\beta)$ gives therefore a bound on the *generalization rate* of the scenario approach, which relates to the ability of the scenario solution of being feasible (with high probability) also with respect to constraints that were not explicitly taken into account in the solution of RCP_N (unseen scenarios). In formula (3), the suffix 'lin' underlines the fact that N grows linearly with respect to β^{-1}.

B. Objective of this paper

In this paper we show that a better bound than (3) in fact holds for scenario convex problems. The new bound (Theorem 2 below) has both theoretical and practical importance. From the theoretical side, it shows that generalization is achieved with a number of samples that grows essentially as $O(\frac{n_\theta}{\varepsilon}\ln\frac{1}{\beta})$. This implies that a much lower number of constraints is needed with respect to (3), which is important in practice when solving RCP_N numerically.

C. Related works

The idea of pursuing robustness in a probabilistic sense is not new, but its use for robust control synthesis is relatively recent. We direct the reader to the recent monograph [23] for an historical perspective on the topic and for a thorough survey of currently available randomized algorithms for approximately solving probabilistically constrained design problems in control.

However, the randomized approach that we propose in this paper is distinctively different from those discussed in [23] and in other related works such as [9], [14], [20], [21], [22]. These latter references propose sequential stochastic algorithms for determining an approximately feasible design, based on random gradient descent or ellipsoidal iterations. For space reasons, we shall not discuss further here these methods, but direct the reader to [23] and to the introduction in [6].

II. Main Result

We start with a simplifying assumption that is made in order to avoid mathematical cluttering.

Assumption 1: For all possible extractions $\delta^{(1)},\ldots,\delta^{(N)}$, the optimization problem (2) is either unfeasible, or, if feasible, it attains a unique optimal solution. ⋆

This assumption could actually be removed (i.e. we may allow for non-existence or non-uniqueness of the optimal solution) without harming the result, at the expense of complications in the proofs.

We now state the main result of this paper.

Theorem 2: Let Assumption 1 be satisfied. Fix two real numbers $\varepsilon\in(0,1)$ (level parameter) and $\beta\in(0,1)$ (confidence parameter). If

$$N\geq N_{\mathrm{gen}}(\varepsilon,\beta)\doteq \tag{4}$$
$$\left\lceil\inf_{\nu\in(0,1)}\frac{1}{1-\nu}\left(\frac{1}{\varepsilon}\ln\frac{1}{\beta}+n_\theta+\frac{n_\theta}{\varepsilon}\ln\frac{1}{\nu\varepsilon}+\frac{1}{\varepsilon}\ln\frac{(n_\theta/\mathrm{e})^{n_\theta}}{n_\theta!}\right)\right\rceil$$

($\lceil\rceil$ denotes the smallest integer greater or equal than the argument) then, with probability no smaller than $1-\beta$, either the scenario problem RCP_N is unfeasible, and hence also RCP is unfeasible; or, RCP_N is feasible, and then its optimal solution $\hat{\theta}_N$ is ε-level robustly feasible. ⋆

A proof for Theorem 2 is presented in Section III. In the theorem, probability $1-\beta$ refers to the probability Prob^N ($=\mathrm{Prob}\times\cdots\times\mathrm{Prob}$, N times) of extracting a 'bad' multisample, i.e. a multisample $\delta^{(1)},\ldots,\delta^{(N)}$ such that $\hat{\theta}_N$ does not meet the ε-level feasibility property. In other words, Theorem 2 states that if N (specified by (4)) random scenarios are drawn, the optimal solution of RCP_N is ε-level feasible according to Definition 2, with high probability $1-\beta$. Bound (4) can be simplified and made explicit, as stated in the following corollary.

Corollary 1: The results in Theorem 2 hold for

$$N\geq N_{\log}(\varepsilon,\beta)\doteq\left\lceil\frac{2}{\varepsilon}\ln\frac{1}{\beta}+2n_\theta+\frac{2n_\theta}{\varepsilon}\ln\frac{2}{\varepsilon}\right\rceil. \tag{5}$$

⋆

Proof: Observe that $(n_\theta/\mathrm{e})^{n_\theta}\leq n_\theta!$, and hence the last term in (4) is non-positive and can be dropped, leading to

$$N_{\mathrm{gen}}(\varepsilon,\beta)\leq\left\lceil\frac{1}{1-\nu}\left(\frac{1}{\varepsilon}\ln\frac{1}{\beta}+n_\theta+\frac{n_\theta}{\varepsilon}\ln\frac{1}{\nu\varepsilon}\right)\right\rceil, \tag{6}$$

where ν can be freely selected in $(0,1)$. The statement of the corollary is then obtained by selecting $\nu=1/2$ in (6). We also note that further optimizing (6) with respect to ν always leads to a $\nu\leq 1/2$, with a corresponding improvement by at most of a factor 2. ■

	$\varepsilon = 0.1$	$\varepsilon = 0.01$	$\varepsilon = 0.001$	$\varepsilon = 0.0001$
$\beta = 0.01$	$N_{\text{lin}} = 10^4$ $N_{\text{log}} = 712$ $N_{\text{gen}} = 533$	$N_{\text{lin}} = 10^5$ $N_{\text{log}} = 11538$ $N_{\text{gen}} = 7940$	$N_{\text{lin}} = 10^6$ $N_{\text{log}} = 161249$ $N_{\text{gen}} = 105142$	$N_{\text{lin}} = 10^7$ $N_{\text{log}} = 2072821$ $N_{\text{gen}} = 1303039$
$\beta = 0.001$	$N_{\text{lin}} = 10^5$ $N_{\text{log}} = 758$ $N_{\text{gen}} = 562$	$N_{\text{lin}} = 10^6$ $N_{\text{log}} = 11999$ $N_{\text{gen}} = 8203$	$N_{\text{lin}} = 10^7$ $N_{\text{log}} = 165854$ $N_{\text{gen}} = 107683$	$N_{\text{lin}} = 10^8$ $N_{\text{log}} = 2118873$ $N_{\text{gen}} = 1327959$
$\beta = 0.0001$	$N_{\text{lin}} = 10^6$ $N_{\text{log}} = 804$ $N_{\text{gen}} = 589$	$N_{\text{lin}} = 10^7$ $N_{\text{log}} = 12459$ $N_{\text{gen}} = 8465$	$N_{\text{lin}} = 10^8$ $N_{\text{log}} = 170459$ $N_{\text{gen}} = 110219$	$N_{\text{lin}} = 10^9$ $N_{\text{log}} = 2164925$ $N_{\text{gen}} = 1352842$
$\beta = 0.00001$	$N_{\text{lin}} = 10^7$ $N_{\text{log}} = 850$ $N_{\text{gen}} = 617$	$N_{\text{lin}} = 10^8$ $N_{\text{log}} = 12920$ $N_{\text{gen}} = 8725$	$N_{\text{lin}} = 10^9$ $N_{\text{log}} = 175064$ $N_{\text{gen}} = 112748$	$N_{\text{lin}} = 10^{10}$ $N_{\text{log}} = 2210977$ $N_{\text{gen}} = 1377687$

TABLE I

COMPARISON OF SAMPLE-SIZE BOUNDS, FOR $n_\theta = 10$.

Remark 1 (sample complexity): Notice that bound (5) — and hence (4), which is tighter — substantially improves upon (3) in that dependence on $1/\beta$ is now logarithmic. Table I shows a comparison of the these bounds for several values of ε and β. ⋆

Remark 2 (the role of convexity): Theorem 2 says that if we extract a *finite* number N of constraints, then the solution of the randomized problem, if feasible, satisfies most of the other unseen constraints. As we mentioned, this is a *generalization* property: the explicit satisfaction of some 'training' scenarios generalizes automatically to the satisfaction of other unseen scenarios. It is interesting to note that generalization calls for some kind of structure, and the only structure used here is convexity. So, convexity in the scenario approach is fundamental in two different respects: on the computational side, it allows for an efficient solution of the ensuing optimization problem, and on the theoretical side it allows for generalization. ⋆

Remark 3 (VC-dimension): Bound (5) depends on the problem structure through n_θ, the number of optimization variables, only. It is not difficult to conceive situations where the class of sets $\{\delta \in \Delta : f(\theta,\delta) > 0\} \subseteq \Delta$, parameterized in θ, has infinite VC-dimension (see e.g. [24] for a definition), even for small n_θ. Then, estimating $\text{Prob}\{\delta \in \Delta : f(\theta,\delta) > 0\} = V(\theta)$ uniformly with respect to θ is impossible and the VC-theory is of no use. Theorem 2 says that, if attention is restricted to $\hat\theta_N$, then estimating $V(\hat\theta_N)$ becomes possible at a low computational cost. ⋆

III. PRELIMINARIES AND PROOF OF THEOREM 2

A. Preliminiaries

We first state a key instrumental result. Consider the convex optimization program

$$\mathscr{P}: \quad \min_{x\in\mathbb{R}^n} c^T x \quad \text{subject to: } x \in \bigcap_{i\in\{1,\ldots,m\}} \mathscr{X}_i,$$

where $\mathscr{X}_i$, $i = 1,\ldots,m$, are closed convex sets, and define the convex programs $\mathscr{P}_k$, $k = 1,\ldots,m$, obtained from $\mathscr{P}$ by removing the k-th constraint:

$$\mathscr{P}_k: \quad \min_{x\in\mathbb{R}^n} c^T x \quad \text{subject to: } x \in \bigcap_{i\in\{1,\ldots,m\}\setminus k} \mathscr{X}_i.$$

Let x^* be any optimal solution of $\mathscr{P}$ (assuming it exists), and let x_k^* be any optimal solution of $\mathscr{P}_k$ (again, assuming it exists). We have the following definition.

Definition 3 (support constraint): The k-th constraint $\mathscr{X}_k$ is a *support constraint* for $\mathscr{P}$ if problem $\mathscr{P}_k$ has an optimal solution x_k^* such that $c^T x_k^* < c^T x^*$. ⋆

The following theorem holds.

Theorem 3 (Theorem 2 in [7]): The number of support constraints for problem $\mathscr{P}$ is at most n. ⋆

B. Proof of Theorem 2

The fact that, if RCP_N is unfeasible, then RCP is unfeasible too is trivially true, since RCP exhibits more constraints than RCP_N. Thus, we have to prove that, with probability $1-\beta$, either RCP_N is unfeasible or, if feasible, its solution is ε-level robustly feasible. This part of the proof is inspired by a similar derivation given in a different context in [13].

For clarity of exposition, we first assume that problem RCP_N is feasible for any selection of $\delta^{(1)}, \ldots, \delta^{(N)}$. The case where infeasibility can occur is obtained as an easy extension as indicated at the end of the proof.

Given N scenarios $\delta^{(1)},\ldots,\delta^{(N)}$, select a subset $I = \{i_1,\ldots,i_{n_\theta}\}$ of n_θ indices from $\{1,\ldots,N\}$ and let $\hat\theta_I$ be the optimal solution of the program

$$\min_{\theta\in\Theta} c^T\theta \quad \text{subject to: } f(\theta,\delta^{(i_j)}) \le 0,\ j = 1,\ldots,n_\theta.$$

Based on $\hat\theta_I$ we next introduce a subset Δ_I^N of the set Δ^N defined as

$$\Delta_I^N \doteq \{(\delta^{(1)},\ldots,\delta^{(N)}) : \quad \hat\theta_I = \hat\theta_N\} \tag{7}$$

($\hat{\theta}_N$ is the optimal solution with all N constraints $\delta^{(1)},\dots,\delta^{(N)}$ in place).

Let now I range over the collection $\mathscr{I}$ of all possible choices of n_θ indices from $\{1,\dots,N\}$ ($\mathscr{I}$ contains $\left(\begin{array}{c} N \\ n_\theta \end{array}\right)$ sets). We want to prove that

$$\Delta^N = \bigcup_{I\in\mathscr{I}} \Delta_I^N. \tag{8}$$

To show (8), take any $(\delta^{(1)},\dots,\delta^{(N)}) \in \Delta^N$. From the set of constraint $\delta^{(1)},\dots,\delta^{(N)}$ eliminate a constraint which is not a support constraint (this is possible in view of Theorem 3, since $N > n_\theta$). The resulting optimization problem with $N-1$ constraints admits the same optimal solution $\hat{\theta}_N$ as the original problem with N constraints. Consider now the set of the remaining $N-1$ constraints and, among these, remove a constraint which is not a support constraint for the problem with $N-1$ constraints. Again, the optimal solution does not change. If we keep going this way until we are left with n_θ constraints, in the end we still have $\hat{\theta}_N$ as optimal solution, showing that $(\delta^{(1)},\dots,\delta^{(N)}) \in \Delta_I^N$, where I is the set containing the n_θ constraints remaining at the end of the process. Since this is true for any choice of $(\delta^{(1)},\dots,\delta^{(N)}) \in \Delta^N$, (8) is proven.

Next, let

$$B \doteq \{(\delta^{(1)},\dots,\delta^{(N)}) : \quad V(\hat{\theta}_N) > \varepsilon\}$$

and

$$B_I \doteq \{(\delta^{(1)},\dots,\delta^{(N)}) : \quad V(\hat{\theta}_I) > \varepsilon\}$$

We now have:

$$\begin{aligned} B &= B\cap\Delta^N \\ &= B\cap(\cup_{I\in\mathscr{I}}\Delta_I^N) \quad \text{(apply (8))} \\ &= \cup_{I\in\mathscr{I}}(B\cap\Delta_I^N) \\ &= \cup_{I\in\mathscr{I}}(B_I\cap\Delta_I^N). \quad \text{(because of (7))} \end{aligned} \tag{9}$$

A bound for $\mathrm{Prob}^N(B)$ is now obtained by bounding $\mathrm{Prob}(B_I\cap\Delta_I^N)$ and then summing over $I\in\mathscr{I}$.

Fix any I, e.g. $I=\{1,\dots,n_\theta\}$ to be more explicit. The set $B_I = B_{\{1,\dots,n_\theta\}}$ is in fact a cylinder with base in the cartesian product of the first n_θ constraint domains (this follows from the fact that condition $V(\hat{\theta}_{\{1,\dots,n_\theta\}}) > \varepsilon$ only involves the first n_θ constraints). Fix $(\bar{\delta}^{(1)},\dots,\bar{\delta}^{(n_\theta)}) \in$ base of the cylinder. For a point $(\bar{\delta}^{(1)},\dots,\bar{\delta}^{(n_\theta)},\delta^{(n_\theta+1)},\dots,\delta^{(N)})$ to be in $B_{\{1,\dots,n_\theta\}}\cap\Delta_{\{1,\dots,n_\theta\}}^N$, constraints $\delta^{(n_\theta+1)},\dots,\delta^{(N)}$ must be satisfied by $\hat{\theta}_{\{1,\dots,n_\theta\}}$, for, otherwise, we would not have $\hat{\theta}_{\{1,\dots,n_\theta\}} = \hat{\theta}_N$, as it is required in $\Delta_{\{1,\dots,n_\theta\}}^N$. But, $V(\hat{\theta}_{\{1,\dots,n_\theta\}}) > \varepsilon$ in $B_{\{1,\dots,n_\theta\}}$. Thus, by the fact that the extractions are independent, we conclude that

$$\begin{aligned} &\mathrm{Prob}^{N-n_\theta}\{(\delta^{(n_\theta+1)},\dots,\delta^{(N)}) : \\ &\quad (\bar{\delta}^{(1)},\dots,\bar{\delta}^{(n_\theta)},\delta^{(n_\theta+1)},\dots,\delta^{(N)}) \in B_{\{1,\dots,n_\theta\}}\cap\Delta_{\{1,\dots,n_\theta\}}^N\} \\ &\qquad < (1-\varepsilon)^{N-n_\theta}. \end{aligned}$$

The probability on the left hand side is nothing but the conditional probability that $(\delta^{(1)},\dots,\delta^{(N)}) \in B_{\{1,\dots,n_\theta\}}\cap\Delta_{\{1,\dots,n_\theta\}}^N$ given $\delta^{(1)} = \bar{\delta}^{(1)},\dots,\delta^{(n_\theta)} = \bar{\delta}^{(n_\theta)}$. Integrating over the base of the cylinder $B_{\{1,\dots,n_\theta\}}$, we then obtain

$$\begin{aligned} &\mathrm{Prob}^N(B_{\{1,\dots,n_\theta\}}\cap\Delta_{\{1,\dots,n_\theta\}}^N) \\ &\quad < \quad (1-\varepsilon)^{N-n_\theta}\cdot\mathrm{Prob}^{n_\theta}(\text{base of } B_{\{1,\dots,n_\theta\}}) \\ &\quad \le \quad (1-\varepsilon)^{N-n_\theta}. \end{aligned}$$

From (9), we finally arrive to the desired bound for $\mathrm{Prob}^N(B)$

$$\mathrm{Prob}^N(B) \le \sum_{I\in\mathscr{I}} \mathrm{Prob}^N(B_I\cap\Delta_I) < \left(\begin{array}{c} N \\ n_\theta \end{array}\right)(1-\varepsilon)^{N-n_\theta}. \tag{10}$$

The last part of the proof is nothing but algebraic manipulations on bound (10) to show that, if N is chosen according to (4), then

$$\left(\begin{array}{c} N \\ n_\theta \end{array}\right)(1-\varepsilon)^{N-n_\theta} \le \beta, \tag{11}$$

so concluding the proof. These manipulations are reported next.

Any of the following inequality implies the next in a top-down fashion, where the first one is (4):

$$\begin{aligned} &N \ge \\ &\quad \tfrac{1}{1-\nu}\left(\tfrac{1}{\varepsilon}\ln\tfrac{1}{\beta} + n_\theta + \tfrac{n_\theta}{\varepsilon}\ln\tfrac{1}{\nu\varepsilon} + \tfrac{1}{\varepsilon}\ln\left(\left(\tfrac{n_\theta}{e}\right)^{n_\theta}\tfrac{1}{n_\theta!}\right)\right); \\ &(1-\nu)N \ge \\ &\quad \tfrac{1}{\varepsilon}\ln\tfrac{1}{\beta} + n_\theta + \tfrac{n_\theta}{\varepsilon}\ln\tfrac{1}{\nu\varepsilon} + \tfrac{1}{\varepsilon}\ln\left(\left(\tfrac{n_\theta}{e}\right)^{n_\theta}\tfrac{1}{n_\theta!}\right); \\ &(1-\nu)N \ge \\ &\quad \tfrac{1}{\varepsilon}\ln\tfrac{1}{\beta} + n_\theta + \tfrac{n_\theta}{\varepsilon}\left(\ln\tfrac{n_\theta}{\nu\varepsilon} - 1\right) - \tfrac{1}{\varepsilon}\ln(n_\theta!); \\ &N \ge \\ &\quad \tfrac{1}{\varepsilon}\ln\tfrac{1}{\beta} + n_\theta + \tfrac{n_\theta}{\varepsilon}\left(\ln\tfrac{n_\theta}{\nu\varepsilon} - 1 + \tfrac{\nu N\varepsilon}{n_\theta}\right) - \tfrac{1}{\varepsilon}\ln(n_\theta!); \\ &N \ge \frac{1}{\varepsilon}\ln\frac{1}{\beta} + n_\theta + \frac{n_\theta}{\varepsilon}\ln N - \frac{1}{\varepsilon}\ln(n_\theta!) \end{aligned} \tag{12}$$

where the last implication can be justified by observing that $\ln x \ge 1-\frac{1}{x}$, for $x>0$, and applying this inequality with $x = \frac{n_\theta}{\gamma N\varepsilon}$. Proceeding from (12), the next inequalities in the chain are

$$\begin{aligned} \ln\beta &\ge -\varepsilon N + \varepsilon n_\theta + n_\theta\ln N - \ln(n_\theta!) \\ \beta &\ge \frac{N^{n_\theta}}{n_\theta!}e^{-\varepsilon(N-n_\theta)} \\ \beta &\ge \frac{N(N-1)\cdot\dots\cdot(N-n_\theta+1)}{n_\theta!}(1-\varepsilon)^{N-n_\theta}, \end{aligned}$$

where, in the last implication, we have used the fact that $e^{-\varepsilon(N-n_\theta)} \ge (1-\varepsilon)^{N-n_\theta}$, as it follows by taking logarithm of the two sides and further noting that $-\varepsilon \ge \ln(1-\varepsilon)$. Finally, we have

$$\beta \ge \left(\begin{array}{c} N \\ n_\theta \end{array}\right)(1-\varepsilon)^{N-n_\theta},$$

which is (11).

So far, we have assumed that RCP_N is feasible for any selection of $\delta^{(1)},\dots,\delta^{(N)}$. Relax now this assumption and call $F \subseteq \Delta^N$ the set where RCP_N is indeed feasible. The same derivation can then be worked out in the domain F, instead of Δ^N, leading to the conclusion that (10) holds with $B \doteq \left\{(\delta^{(1)},\ldots,\delta^{(N)}) \in F : V(\hat{\theta}_N) > \varepsilon\right\}$, which concludes the proof.

■

IV. Example: Robust State-Feedback Stabilization

Given the uncertain system

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

we wish to design a control $u = Kx$ such that the closed-loop is quadratically stable, for all δ in the allowable uncertainty set Δ. This design specification is satisfied if and only if there exist $P \succ 0$ (means P is symmetric and positive-definite) and Y such that

$$A(\delta)P + PA^T(\delta) + BY + Y^T B \prec 0, \quad \forall \delta \in \Delta.$$

Due to homogeneity in these conditions, we can reformulate the problem in minimization form as the RCP

$$\min_{P,Y,\nu} \nu \text{ subject to}$$
$$-I \preceq A(\delta)P + PA^T(\delta) + BY + Y^T B \preceq \nu I, \ \forall \delta \in \Delta$$
$$I \succeq P \succeq -\nu I.$$

If optimal ν is negative, then the original design conditions are satisfied, and the controller is retrieved as $K = YP^{-1}$.

We here consider a simple numerical example, with

$$A(\delta) = \begin{bmatrix} \rho_2\delta_2 & 1+\rho_1\delta_1 \\ -(1+\rho_1\delta_1)^2 & 2(0.1+\rho_2\delta_2)(1+\rho_1\delta_1) \end{bmatrix}$$
$$B = \begin{bmatrix} 10 \\ 15 \end{bmatrix}$$

$\rho_1 = 1$, $\rho_2 = 0.5$, with $|\delta_1| \le 1$, $|\delta_2| \le 1$. The scenario counterpart of the problem is

$$\min \nu \text{ subject to}$$
$$-I \preceq A(\delta^{(i)})P + PA^T(\delta^{(i)}) + BY + Y^T B \preceq \nu I, \ i = 1,\ldots,N$$
$$I \succeq P \succeq -\nu I,$$

where $\delta^{(1)},\ldots,\delta^{(N)}$ are iid uncertainty samples.

In this example we have $n_\theta = 3+2+1 = 6$ design variables (the free entries of symmetric P, plus the two entries of Y, and ν). Setting $\varepsilon = 0.01$ and $\beta = 0.001$, bound (4) requires at least $N = 5170$ uncertainty samples. We then selected the uniform probability measure over Δ, and solved numerically (by means of LMILab toolbox in Matlab) one instance of the scenario problem. This yielded the optimal solution $\alpha = -3.1065 \times 10^{-5}$ and

$$P = \begin{bmatrix} 0.0000516 & 0.0000401 \\ 0.0000401 & 0.2885159 \end{bmatrix},$$
$$Y = \begin{bmatrix} -0.0123178 & -0.0183152 \end{bmatrix}$$

and hence the controller

$$K = YP^{-1} = [-238.70813 \ \ -0.0302966].$$

This controller was then tested a-posteriori with a Monte Carlo analysis. The a-posteriori estimated probability of P,Y violating the original design LMIs was 1.15×10^{-5} (estimated using 6×10^6 uniform samples). This means in practice that the computed P is a Lyapunov matrix for all but a very small fraction of the closed-loop plants.

V. Conclusions

Efficient and exact solution methodologies for robust convex problems are known only for certain simple dependencies of the constraints on the uncertainty (e.g. affine, polynomial or rational). In all other cases, the scenario approach offers a viable route to find a 'solution' to the design problem.

Even when solving RCP is possible, resorting to the scenario approach can be advantageous because it alleviates the conservativeness inherent in RCP. In fact, solving RCP gives a 100% deterministic guarantee that the constraints are satisfied. On the other hand, accepting a small risk of constraint violation can result in a (sometimes significant) performance improvement for all plants whose constraints are satisfied. In this connection, ε can be seen as a 'tuning-knob' that permits to trade probability of unfeasibility for performance.

The sample complexity of scenario optimization has been drastically reduced by the new bound derived in this paper, thus hopefully increasing the appeal of this technique for application to practical engineering design problems.

References

[1] P. Apkarian and H.D. Tuan. Parameterized LMIs in control theory. *SIAM Journal on Control and Optimization*, 38(4):1241–1264, 2000.

[2] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semidefinite programming. *SIAM J. on Optimization*, 7(4):991–1016, 1997.

[3] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[4] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[5] A. Ben-Tal and A. Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12(3):811–833, 2002.

[6] G. Calafiore and M.C. Campi. Robust convex programs: randomized solutions and confidence levels. In *Proc. IEEE Conf. on Decision and Control*, Maui, USA, December 2003.

[7] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming – Online First*, DOI 10.1007/s10107-003-0499-y, 2004.

[8] G. Calafiore and L. El Ghaoui. Ellipsoidal bounds for uncertain linear equations and dynamical systems. *Automatica*, 50(5):773–787, 2004.

[9] G. Calafiore and B.T. Polyak. Fast algorithms for exact and approximate feasibility of robust LMIs. *IEEE Trans. Aut. Control*, 46(11):1755–1759, November 2001.

[10] M.C. de Oliveira, J. Bernussou, and J.C. Geromel. A new discrete-time robust stability condition. *Systems and Control Letters*, 37:261–265, 1999.

[11] M.C. de Oliveira, J.C. Geromel, and J. Bernussou. Extended H_2 and H_∞ norm characterization and controller parameterization for discrete-time systems. *Int. J. Control*, 75(9):666–679, 2002.

[12] E. Feron, P. Apkarian, and P. Gahinet. Analysis and synthesis of robust control systems via parameter-dependent Lyapunov functions. *IEEE Trans. Aut. Control*, 41(7):1041–1046, 1996.

[13] S. Floyd and M. Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, pages 1–36, 1995.

[14] Y. Fujisaki, F. Dabbene, and R. Tempo. Probabilistic design of LPV control systems. *Automatica*, 39(8):1323–1337, 2003.

[15] J.C. Geromel, J. Bernussou, G. Garcia, and M.C. de Oliveira. H_2 and H_1 robust filtering for discrete-time linear systems. *SIAM Journal on Control and Optimization*, 38(5):1353–1368, 2000.

[16] L. El Ghaoui and G. Calafiore. Robust filtering for discrete-time systems with bounded noise and parametric uncertainty. *IEEE Trans. Aut. Control*, 46(7):1084–1089, July 2001.

[17] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM J. on Matrix Analysis and Applications*, 18(4):1035–1064, 1997.

[18] L. El Ghaoui and H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.*, 9(1):33–52, 1998.

[19] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, July 2003.

[20] S. Kanev, B. De Schutter, and M. Verhaegen. An ellipsoid algorithm for probabilistic robust controller design. *Systems and Control Letters*, 49(5):365–375, 2003.

[21] Y. Oishi and H. Kimura. Randomized algorithms to solve parameter-dependent linear matrix inequalities and their computational complexity. In *Proc. IEEE Conf. on Decision and Control*, volume 2, pages 2025–2030, December 2001.

[22] B.T. Polyak and R. Tempo. Probabilistic robust design with linear quadratic regulators. *Systems and Control Letters*, 43:343–353, 2001.

[23] R. Tempo, G. Calafiore, and F. Dabbene. *Randomized Algorithms for Analysis and Control of Uncertain Systems*. Communications and Control Engineering Series. Springer-Verlag, London, 2004.

[24] V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1996.

[25] L. Xie, Y.C. Soh, and C.E. de Souza. Robust Kalman filtering for uncertain discrete-time systems. *IEEE Trans. Aut. Control*, 39(6):1310–1314, 1994.