A scenario approach to non-convex control design:
preliminary probabilistic guarantees

Sergio Grammatico, Xiaojing Zhang, Kostas Margellos, Paul Goulart and John Lygeros

Abstract—Randomized optimization is a recently established tool for control design with modulated robustness. While for uncertain convex programs there exist randomized approaches with efficient sampling, this is not the case for non-convex problems. Approaches based on statistical learning theory are applicable for a certain class of non-convex problems, but they usually are conservative in terms of performance and are computationally demanding.

In this paper, we present a novel scenario approach for a wide class of random non-convex programs. We provide a sample complexity similar to the one for uncertain convex programs, but valid for all feasible solutions inside a set of a-priori chosen complexity. Our scenario approach applies to many non-convex control-design problems, for instance control synthesis based on uncertain bilinear matrix inequalities.

I. INTRODUCTION

Modern control design often relies on the solution of an optimization problem, for instance in Model Predictive Control (MPC) [1] and Lyapunov-based optimal control [2]. In almost all practical control applications, the data describing the plant dynamics are uncertain. The classic way of dealing with the uncertainty is the robust, also called “min-max” or “worst-case”, approach in which the control design has to satisfy the given specifications for all possible realizations of the uncertainty. The worst-case approach is often formulated as a robust optimization problem. However, even robust convex programs are not easy to solve [3]. In addition, from an engineering perspective, robust solutions tend to be conservative in terms of performance.

In order to reduce the conservatism of robust solutions, stochastic programming [4], [5] offers an alternative paradigm. Unlike the worst-case approach, the constraints of the problem can be treated in a probabilistic sense via chance constraints [6], allowing for constraint violations with chosen low probability. The main issue of Chance Constrained Programs (CCPs) is that, without assumptions on the underlying probability distribution, they are in general intractable because multi-dimensional probability integrals must be computed.

Among the class of chance constrained programs, “uncertain convex programs” have received particular attention [7], although the feasible set of an uncertain convex programs is in general nonconvex. An established and computationally-tractable approach to approximate chance constrained problems is the scenario approximation [7]. A solution to the CCP is found with high confidence by solving an optimization problem, called Scenario Program (SP), subject to a finite number of randomly drawn constraints (scenarios). This scenario approach is particularly effective whenever it is possible to generate samples from the uncertainty, without any further knowledge on its probabilistic nature. From a practical point of view, this is generally the case for many control-design problems where historical data and/or predictions are available. The scenario approach for general uncertain convex programs was first introduced in [8], and many control-design applications are outlined in [9]. The fundamental contribution in these works is the explicit characterization of the number of scenarios (“sample complexity”) needed to guarantee that the optimal solution of the SP is a feasible solution to the original CCP with high confidence. The sample complexity is then further refined in [10] where it is shown to be tight for “fully-supported” problems, in [11] where the concept of Helly’s dimension is introduced to reduce the conservatism for non-fully supported problems; moreover, refinements based on the structure of the constraints are presented in [12] and [13], [14].

While feasibility, optimality and sample complexity of random convex programs are well characterized, to the best of the authors’ knowledge, there exists no scenario approach to characterize random non-convex programs. One way to deal with these problems comes from statistical learning theory, based on the Vapnik-Chervonenkis (VC) theory [15], [16], [17], and it can be applied to many non-convex control-design problems [18], [19]. The sample complexity from statistical learning theory provides probabilistic guarantees for all feasible solutions of the sampled program and not just for the optimal solution, contrary to the results in [10], [11]. This distinction is fundamental because the global optimizer of non-convex programs is in general not computable, so it is necessary to provide probabilistic guarantees for all solutions in a feasible set. However, the more-general probabilistic guarantees of VC theory come at the price of a quite large number of random samples [9]. More fundamentally, they depend on the so-called VC dimension which is in general difficult to compute, or even infinite, in which case the VC theory is not even applicable [8].

The aim of this paper is to present a scenario approach for a wide class of random non-convex programs, which
comes with efficient sample complexity, and provides probabilistic guarantees for all feasible solutions in a set of a-priori chosen complexity. In the spirit of [8], [9], [10], [11], our results are only based on the decision complexity, while no assumption is made on the underlying probability structure. We present a scenario approach for the class of random non-convex programs with (possibly) non-convex cost, deterministic (possibly) non-convex constraints, and chance constraint containing functions with separable non-convexity; for this class of programs, Helly’s dimension (associated with the global optimal value) can be unbounded [20]. This means that the standard scenario approach is not directly applicable, hence motivating our methodology. We provide a sample complexity $O(n/\epsilon (n + \ln (M/\beta)))$ for all feasible solutions inside a convex set, where $n$ is the decision dimension, $\epsilon, \beta$ are the usual desired levels of probabilistic feasibility, and the integer $M$ denotes a desired “degree of complexity” of the derived feasibility set.

The paper is structured as follows. Section II presents the technical background and the problem statement. Section III presents the main results. Further discussion and comparisons are given in Section IV. Section V presents a scenario approach for control design via uncertain Bilinear Matrix Inequalities (BMIs). We conclude the paper in Section VI.

Notation

$\mathbb{R}$ and $\mathbb{Z}$ denote, respectively, the set of real and integer numbers. The notation $[a, b]$ denotes the integer interval $\{a, a + 1, \ldots, b\} \subseteq \mathbb{Z}$. Given a square matrix $A \in \mathbb{R}^{n \times n}$, $\lambda_{\text{max}}(A)$ denotes its maximum eigenvalue. The notation $\text{conv} (\cdot)$ denotes the convex hull. We use the short-hand notation $\mathbb{P} \{g(x, \cdot) \leq 0\}$ in place of $\mathbb{P} \{\{\delta \in \Delta \mid g(x, \delta) \leq 0\}\}$; analogously, we use $\mathbb{P}^N \{\{\omega \in \Delta^N \mid V(X(\omega)) > \epsilon\}\}$.

II. CHANCE CONSTRAINED PROGRAMMING

We consider a Chance Constrained Program (CCP) with cost function $J : \mathbb{R}^n \to \mathbb{R}$, constraint function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, constraint-violation tolerance $\epsilon \in (0, 1)$, and admissible set $\mathcal{X} \subseteq \mathbb{R}^n$.

$$\text{CCP}(\epsilon) : \begin{cases} \min_{x \in \mathcal{X}} J(x) \\ \text{sub. to: } \mathbb{P} \{g(x, \cdot) \leq 0\} \geq 1 - \epsilon. \end{cases} \tag{1}$$

We assume that $\text{CCP}(\epsilon)$ in (1) is feasible. The random variable in (1) is $\delta$, defined on a probability space $(\Delta, \mathcal{F}, \mathbb{P})$, with $\Delta \subseteq \mathbb{R}^m$. Measure-theoretic details are given in [20].

Throughout the paper, we consider the following assumption [11, Assumption 1].

Standing Assumption 1: The set $\mathcal{X} \subseteq \mathbb{R}^n$ is compact and convex$^1$. For all $\delta \in \Delta \subseteq \mathbb{R}^m$, the mapping $x \mapsto g(x, \delta)$ is convex and lower semicontinuous.

The compactness assumption, typical of any problem of practical interest, avoids technical difficulties by guaranteeing that any feasible problem instance attains an optimal solution [11, Section 3.1, pag. 3433].

It is important to notice that unlike the standard setting of random convex programs [8], [9], [10], [11], we allow the cost function $J$ to be non-convex. As shown in [20], this immediately implies that Helly’s dimension (associated with the optimizer mapping) can be unbounded, therefore the standard scenario approach is not directly applicable. We also notice that the CCP formulation in (1) implicitly includes the more general CCP$'(\epsilon)$:

$$\begin{cases} \min_{x \in \mathcal{X}} J(x) \\ \text{sub. to: } \mathbb{P} \{\lambda (g(x, \cdot) + f(x)\varphi(\cdot)) \leq 0\} \geq 1 - \epsilon \\ h(x) \leq 0, \end{cases} \tag{2}$$

for possibly non-convex functions $f, h, \varphi$ and convex function $\lambda$. This can be shown introducing an extra variable $y = f(x)$ and then following the lines of [21, Section 1.A, pag. 6-7].

Since our results rely on probabilistic guarantees for an entire set, rather than for a single point, we define set-based counterparts of [9, Definitions 1, 2].

Definition 1 (Probability of Violation of a Set): For any set $\mathcal{X} \subseteq \mathcal{X}'$, the probability of violation of $\mathcal{X}$ is defined as

$$V(\mathcal{X}) := \sup_{x \in \mathcal{X}} \mathbb{P} (\{\delta \in \Delta \mid g(x, \delta) > 0\}). \tag{3}$$

Definition 2 (Feasibility of a Set): For any given $\epsilon \in (0, 1)$, a set $\mathcal{X} \subseteq \mathcal{X}'$ is feasible for CCP$'(\epsilon)$ in (1) if $V(\mathcal{X}) \leq \epsilon$.

In view of Definitions 1, 2, our developments are mainly inspired by the following key result.

Theorem 1: For any $\mathcal{X} \subseteq \mathbb{R}^n$, and $\epsilon \in (0, 1)$, if $V(\mathcal{X}) \leq \epsilon$, then $V(\text{conv}(\mathcal{X})) \leq (n + 1)\epsilon$.

To the best of our knowledge this basic fact has not been observed in the literature. An immediate consequence of Theorem 1 is that the feasibility set $\mathcal{X}_\epsilon := \{x \in \mathcal{X} \mid \mathbb{P} \{g(x, \cdot) \leq 0\} \geq 1 - \epsilon\}$ of CCP$'(\epsilon)$ in (1) satisfies $\mathcal{X}_\epsilon \subseteq \text{conv}(\mathcal{X}_\epsilon) \subseteq \mathcal{X}_{(n+1)\epsilon}$.

Associated to CCP$'(\epsilon)$ in (1), we consider a Scenario Program (SP) obtained from $\mathcal{N}$ independent and identically distributed (i.i.d.) samples $\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$ drawn according to $\mathbb{P}$ [8, Definition 3].

For a fixed multi-sample $\omega := (\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}) \in \Delta^N$, we consider the SP

$$\text{SP}[\omega] : \begin{cases} \min_{x \in \mathcal{X}} J(x) \\ \text{sub. to: } g(x, \delta^{(i)}) \leq 0 \quad \forall i \in \mathbb{Z}[1, N]. \end{cases} \tag{4}$$
Known facts on scenario approximations of chance constraints

According to [10], [11], if \( J(x) = c^T x \), for some cost vectors \( c, \) the optimizer mapping \( x^*() : \Delta^N \rightarrow \mathcal{X} \) of SP[\( \bar{\omega} \)], assuming it is unique [10, Assumption 1], [11, Assumption 2] or a suitable tie-breaking rule is adopted [8, Section 4.1] [10, Section 2.1], is such that

\[
P^N \{ V(\{x^*(\cdot)\}) > \epsilon \} \leq \Phi(\epsilon, n, N) := \sum_{j=0}^{n-1} \binom{N}{j} \epsilon^j (1 - \epsilon)^{N-j}. \tag{5}
\]

The above bound is tight for fully-supported problems [10, Theorem 1, Equation (7)], while for all other problems it can be improved by replacing \( n \) with the so-called Helly’s dimension \( \zeta \) [11, Theorem 3.3], or with the support rank [12, Lemma 3.8]. An explicit bound for the number of samples needed to satisfy \( \Phi(\epsilon, n, N) \leq \beta \) is given by [11, Corollary 5.1]

\[
N \geq \frac{\pi}{\epsilon \zeta \beta} \left( n - 1 + \ln \left( \frac{1}{\beta} \right) \right). \tag{6}
\]

We stress that the inequality (5) holds only for the probability of violation of the singleton mapping \( x^*(\cdot) \).

Although the only explicit difference between the SP in (4) and convex SPs is the non-convex cost \( J \), it is not possible to directly apply the classic scenario approach based on Helly’s theorem as in [8], [9], [10], [11].

The above statement, for which we need to consider a finite number of mappings \( x_1^*, x_2^*, ..., x_M^* : \Delta^N \rightarrow \mathcal{X} \) satisfying the following assumption.

Assumption 1: The mappings \( x_1^*, x_2^*, ..., x_M^* : \Delta^N \rightarrow \mathcal{X} \) are such that, for all \( k \in \mathbb{Z}[1, M] \), and \( \epsilon \in (0, 1) \), we have

\[
P^N \{ V(\{x^*_k(\cdot)\}) > \epsilon \} \leq \beta_k \in (0, 1). \]

Lemma 1: Consider the SP[\( \bar{\omega} \)] in (4) with \( N \geq n \). For any \( \epsilon \in (0, 1) \), if Assumption 1 holds, then

\[
P^N \{ V(\{x^*_1(\cdot), x^*_2(\cdot), ..., x^*_M(\cdot)\}) > \epsilon \} \leq \sum_{k=1}^{M} \beta_k.
\]

In the results in [16, Section 4.2], the decision variable \( x \) lives in a set \( \mathcal{X} \) of finite cardinality. The main difference is that Lemma 1 instead relies on a finite number of mappings \( x^*_k(\cdot) \), each associated with a given upper bound \( \beta_k \) on the probability that \( V(\{x^*_k(\cdot)\}) \) exceeds \( \epsilon \).

We proceed by addressing the CCP(\( \epsilon \)) in (1) through a family of \( M \geq n + 1 \) distinct convex SPs. We consider \( M \) cost vectors \( c_1, c_2, ..., c_M \in \mathbb{R}^n \), chosen arbitrarily. For each \( k \in \mathbb{Z}[1, M] \), we define

\[
SP_k[\bar{\omega}] := \left\{ \min_{x \in \mathcal{X}} c_k^T x \right\} \text{ sub. to: } g(x, \bar{\delta}(i)) \leq 0 \text{ \forall } i \in \mathbb{Z}[1, N]. \tag{8}
\]

For simplicity, we assume that, for all \( k \in \mathbb{Z}[1, M] \), \( SP_k[\bar{\omega}] \) is feasible almost surely. For all \( \omega \in \Delta^N \), let us consider the set

\[
\mathcal{X}_M(\bar{\omega}) := \text{conv} \{ (x_1^*(\bar{\omega}), x_2^*(\bar{\omega}), ..., x_M^*(\bar{\omega})) \}, \tag{9}
\]

where, for all \( k \in \mathbb{Z}[1, M] \), \( x_k^*(\cdot) \) is the unique optimizer mapping of \( SP_k[\bar{\omega}] \) in (8).

After solving the \( M \) SPs from (8) for the given multi-sample \( \bar{\omega} \in \Delta^N \), we can solve the following approximation of CCP(\( \epsilon \)) in (1).

\[
SP[\bar{\omega}] := \left\{ \min_{x \in \mathcal{X}} J(x) \right\} \text{ sub. to: } x \in \mathcal{X}_M(\bar{\omega}). \tag{10}
\]

We are now ready to state our main results. We provide an implicit upper bound on the probability of violating the chance constraint in (1). The distinction with respect to the known inequality in (5), which is valid for convex programs, is that our probabilistic guarantees hold for all feasible solutions to the non-convex program in (10), not just for the optimal solution. Our bound leads to an explicit sample size which is similar to (6), but with scaled \( \epsilon, \beta \). In particular, the degree of complexity of the convex-hull feasibility set defined in (9) only affects the sample size in a logarithmically.

Theorem 2: For each \( k \in \mathbb{Z}[1, M] \), let \( x_k^* \) be the optimizer mapping of \( SP_k[\bar{\omega}] \) in (8), and let \( \mathcal{X}_M \) be as in (9). Then

\[
P^N \{ V(\mathcal{X}_M(\cdot)) > \epsilon \} \leq M \Phi \left( \frac{\epsilon}{\pi \gamma}, n, N \right). \tag{11}
\]

Corollary 1: For each \( k \in \mathbb{Z}[1, M] \), let \( x_k^* \) be the optimizer mapping of \( SP_k[\bar{\omega}] \) in (8), and let \( \mathcal{X}_M \) be as in (9). Let \( \epsilon, \beta \in (0, 1) \). If

\[
N \geq \frac{\pi \gamma (\epsilon + \beta)}{\epsilon} \left( n - 1 + \ln \left( \frac{M}{\beta} \right) \right), \tag{12}
\]

then \( P^N \{ V(\mathcal{X}_M(\cdot)) \leq \epsilon \} \geq 1 - \beta \), i.e., with probability no smaller than \( 1 - \beta \), any feasible solution of \( SP[\bar{\omega}] \) in (10) is feasible for CCP(\( \epsilon \)) in (1).

Note that the feasibility region \( \mathcal{X}_M(\bar{\omega}) \) in (9) is a subset of the feasibility region \( \mathcal{X}(\bar{\omega}) := \{ x \in \mathcal{X} | g(x, \bar{\delta}(i)) \leq 0 \text{ \forall } i \in \mathbb{Z}[1, N] \} \) of \( SP[\bar{\omega}] \) in (4).

2In place of \( SP_k[\bar{\omega}] \) in (8), alternative constructions are presented in [20].
In order to construct $X_M$ so that it is a tighter inner approximation of $X(\bar{\omega})$, the SPs in (8) should be chosen appropriately. The choice in (8) is indeed motivated by the fact that the optimal solution $x_k(\bar{\omega})$ of SP$k(\bar{\omega})$ belongs to the boundary of the feasibility set $X(\bar{\omega})$. The selection of $M$ indeed induces a trade off; however, the influence of $M$ on the sample size $N$ in (12) is merely logarithmic, so large values of $M$ do not substantially increase $N$.

IV. COMPARISON WITH STATISTICAL LEARNING THEORY

Let us consider our sample size in Corollary 1 relative to Statistical Learning methods for random non-convex programs.

In terms of constraint violation tolerance $\epsilon$, our sample size in (12) grows as $1/\epsilon$ while the sample size provided via the classic statistical learning theory, assuming that the sampled programs are feasible almost surely, grows as $1/\epsilon^2$ [23, Sections 4, 5], [19, Chapter 8]. An important refinement of such a sample-size bound is possible considering the so-called one-sided probability of constrained failure, see for instance [16, Chapter 8], [17, Chapter 7], [22, Sections IV, VI], with asymptotic dependence on $\epsilon$ equal to $1/\epsilon \ln(1/\epsilon)$.

Most important, our sample size in (12) only depends on the dimension $n$ of the decision variable, not on the VC dimension $\xi_{VC}$ of the constraint function $g$ which may be difficult to estimate, or even infinite, in which case VC-theory is not applicable.

On the other hand, a disadvantage of the result in Theorem 2 is that the probabilistic guarantees regard any feasible point inside a certain polytopic region, while the probabilistic guarantees provided by standard statistical learning theory regard any feasible point. However, our derived region is the convex hull of $M$ points, and the dependence of the sample size in (12) on $M$ is merely logarithmic. Therefore, we can get a relatively-tight inner approximation of the true (sampled) feasibility set, without substantially affecting the sample size.

Finally, we have to mention that the statistical learning theory approach can potentially address more general non-convex problems, while our results assume that the constraint function has separable non-convexity.

V. APPLICATION TO UNCERTAIN BILINEAR MATRIX INEQUALITIES

The problem of finding the VC dimension of uncertain Bilinear Matrix Inequality (BMI) constraints has been recently solved in [24]. Therefore, the sample size needed to ensure the desired probabilistic guarantees is given by statistical learning theory. Let us indeed consider a BMI problem in the variables $x \in \mathbb{R}^N$, $y \in \mathbb{R}^M$:

$$
\min_{x,y} J(x,y)
$$

subject to:

$$
P \left[ F_0(\cdot) + \sum_{i=1}^{N} \{x_i F_i(\cdot)\} + \sum_{j=1}^{M} \{y_j G_j(\cdot)\} + \sum_{i=1}^{N} \sum_{j=1}^{M} \{x_i y_j H_{i,j}(\cdot)\} \leq 0 \right] \geq 1 - \epsilon.
$$

where $F_0, F_1, ..., F_n, G_1, ..., G_m, H_{1,1}, H_{1,2}, ..., H_{N,M} \in \mathbb{R}^{n \times n}$ are symmetric matrices.

Since the probabilistic constraint in (13) is equivalent to

$$
P \left[ \lambda_{\max} \left( F_0(\cdot) + \sum_{i=1}^{N} \{x_i F_i(\cdot)\} + \sum_{j=1}^{M} \{y_j G_j(\cdot)\} + \sum_{i=1}^{N} \sum_{j=1}^{M} \{x_i y_j H_{i,j}(\cdot)\} \right) \right] \geq 1 - \epsilon,
$$

we can introduce an extra matrix variable $Z \in \mathbb{R}^{N \times M}$ such that $Z_{i,j} = x_i y_j$ for all $i \in [1,N]$, $j \in [1,M]$, in order to achieve convexity (with respect to $(x,y,Z)$) inside the probability constraint, while introducing a non-convex hard constraint. In view of the CCP formulation in (2), this fits in the set-up of Section III. We notice that the (worst-case) size of the new decision variable $(x,y,Z)$ is $N + M + NM$, whenever all the elements of the matrix $H(\delta)$ are uncertain.

On the other hand, if the matrix $H$ is not uncertain, i.e. $H_{i,j}(\delta) = H_{i,j}$ for all indices $i,j$, then we can introduce the extra matrix variable $Z = Z^T = \sum_{i=1}^{N} \sum_{j=1}^{M} x_i y_j H_{i,j} \in \mathbb{R}^{n \times n}$, so that the number of extra decision variables is $n(n+1)/2$, instead of $NM$.

Let us first consider the Static Output Feedback (SOF) control problem. For the sake of simplicity, let only $A = A(\delta) \in \mathbb{R}^{n \times n}$ be uncertain.

$$
\min_{K,P} J(K,P)
$$

subject to:

$$
P \left[ (A(\cdot) + BKC)^T P + P (A(\cdot) + BKC) \leq -\eta I \right] \geq 1 - \epsilon,
$$

where $\eta > 0$ is a given constant, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{n \times p}$. We can hence introduce an extra matrix variable $Z = P BK \in \mathbb{R}^{n \times p}$, so that we get the following non-convex CCP which fits the formulation (2).

$$
\min_{K,P,Z} J(K,P)
$$

subject to:

$$
P \left[ [A(\cdot)^T P + PA(\cdot) + C^T Z^T + ZC] \leq -\eta I \right] \geq 1 - \epsilon
$$

$$
Z = P BK
$$

(15)

The number of decision variables is $n(n+1)/2$ for $P$, $mn$ for $K$, plus $np$ for the additional variable $Z$, therefore $(m+p)n + n(n+1)/2$ in total.

We can compare the required sample size with the one in [24], which presents an upper bound for the VC dimension of (strict) uncertain LMI and BMI programs. In the SOF case, the bound on the VC dimension $\xi_{VC}$ derived in [24, Theorem 3] reads as $\xi_{VC} = 2(n(n+1)/2+mn) \log_2(4en^2n)$, which roughly grows as $n^3 \ln(n)$. Like all methods based on (one-sided) statistical learning theory, the dependence on $\epsilon$ is $1/\epsilon \ln(1/\epsilon)$. Instead, our sample size in (12) grows as $n^4$ and $1/\epsilon$, respectively in terms of state dimension $n$ and tolerance $\epsilon$. As a consequence, our sample size in (12) is slightly worse with respect to the state space dimension $n$, but slightly better in terms of tolerance $\epsilon$.

Let us indeed derive the explicit number of samples needed for the SOF uncertain BMI for the cases: (i) $(x,u,y) \in \mathbb{R}^3 \times \mathbb{R}^2 \times \mathbb{R}$, and (ii) $(x,u,y) \in \mathbb{R}^4 \times \mathbb{R}^2 \times \mathbb{R}^2$. In Table I
we compare our sample size $N$ in (12) with the sample size $N_{\text{VC}}$ in (7) provided by statistical learning theory. The VC dimension $\xi_{\text{VC}}$ in case (i) is upper bounded by 129, while in case (ii) by 265. We notice that for $\epsilon = 0.1, 0.01, 0.001$, we have that $N_{\text{VC}}$ is about 10 times larger than the sample size $N$ required by the proposed approach.

As second example, let us consider the stabilization problem for an uncertain linear system

$$\dot{x} = A(\delta)x + B(\delta)u,$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control input, and $\delta \in \mathbb{R}^d$ is a random uncertainty.

It follows from [25, Theorem 4] that, in case of polytopic uncertainty, the system (16) is robustly stabilizable if and only if the BMI in [25, Equation (29)] is feasible. Informally speaking, the main idea is to consider a control Lyapunov function candidate of the kind $\max_{i \in \mathbb{Z}[1, R]} x^T Q_i^{-1} x$, where $Q_i \succ 0$ for all $i \in \mathbb{Z}[1, R]$ are matrix variables. Alternative, but conceptually similar, choices of the candidate control Lyapunov function are also possible, see [25, 26, 27].

We next address the probabilistic counterpart of the BMI in [25, Equation (29)] as follows.

$$\mathbb{P} \left[ A(\cdot) Q_k + B(\cdot) Y_k + Q_k A(\cdot)^T + Y_k^T B(\cdot)^T \right] \preceq -\eta Q_k + \sum_{j=1}^R \gamma_{j,k} (Q_j - Q_k),$$

$$(17a)$$

$$\gamma_{j,k} \geq 0 \quad \forall j, k \in \mathbb{Z}[1, R].$$

(17b)

We can derive a probabilistic LMI constraint in (17) by introducing $R$ extra matrix variables $Z_k = \sum_{j=1}^R \gamma_{j,k} (Q_j - Q_k) \in \mathbb{R}^{n \times n}$, for $k = 1, 2, ..., R$. In this way, the number of decision variables needed is $S n(n + 1)/2$ for each of the $R$ matrices $Q_1, Q_2, ..., Q_R$, $Rmn$ for the $R$ matrices $K_1, K_2, ..., K_R$, $R^2$ for the matrices $\gamma_{1,1}, \gamma_{1,2}, ..., \gamma_{R,R}$, and $R(n + 1)/2$ for the additional matrices $Z_k$. Therefore we get $Rn(n + 1)/2 + Rmn + R^2 + R(n + 1)/2$ in total.

In view of Corollary 1, with $Rn(n + 1)/2 + Rmn + R^2 + R(n + 1)/2$ decision variables and $M > n$, our sample size grows as $n^4$ with respect to the state space dimension $n$, and as $1/\epsilon$ with respect to the tolerance $\epsilon$. Therefore the comparison between our sample size in (12) and the one recently provided in [24] is qualitatively the same of the SOF example.

### VI. Conclusion

We have considered a scenario approach for the class of random non-convex programs with (possibly) non-convex cost, deterministic (possibly) non-convex constraints, and chance constraint containing functions with separable non-convexity. We have derived an efficient sample complexity for all feasible solutions inside a convex set with chosen complexity, which logarithmically affects the sample size.

Our set-based scenario approach may motivate various non-convex control design applications, for instance randomized MPC of uncertain nonlinear control-affine systems [28].

## Appendix I

### Proofs

**Proof of Theorem 1**

Let $\mathcal{X}_e := \{ x \in \mathcal{X} | \mathbb{P}(\{ \delta \in \Delta \mid g(x, \delta) \leq 0 \}) \geq 1 - \epsilon \}$ be the feasibility set of CCP($\epsilon$) in (1). Take any arbitrary $y \in \mathcal{conv}(\mathcal{X}_e)$. It follows from Caratheodory’s Theorem [29, Theorem 17.1] that there exist $x_1, x_2, ..., x_{n+1} \in \mathcal{X}_e$ such that $y = \sum_{i=1}^{n+1} \alpha_i x_i$ for some $\alpha_1, \alpha_2, ..., \alpha_{n+1} \in [0, 1]$ such that $\sum_{i=1}^{n+1} \alpha_i = 1$.

In the following inequalities, we exploit the convexity of the mapping $x \mapsto g(x, \delta)$ for each fixed $\delta \in \Delta$ from Standing Assumption 1.

$$\mathbb{P} \left[ g(y, \cdot) > 0 \right] \leq \mathbb{P} \left[ g(\sum_{i=1}^{n+1} \alpha_i x_i, \cdot) > 0 \right]$$

$$\leq \sum_{i=1}^{n+1} \alpha_i g(x_i, \cdot) > 0$$

$$\leq \sum_{i=1}^{n+1} \alpha_i g(x_i, \cdot) > 0$$

$$\leq (n + 1)\epsilon.$$  

The last inequality follows from the fact that $x_1, x_2, ..., x_{n+1} \in \mathcal{X}_e$. Since $y \in \mathcal{conv}(\mathcal{X}_e)$ has been chosen arbitrarily, it follows that $V(\mathcal{conv}(\mathcal{X}_e)) \leq (n + 1)\epsilon$.

**Proof of Lemma 1**

$$\mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(\{ x^1_1(\omega), ..., x^M_1(\omega) \}) > \epsilon \} \right)$$

$$\leq \sum_{k=1}^M \mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(\{ x^k_1(\omega) \}) > \epsilon \} \right)$$

where the last inequality follows from Assumption 1.

**Proof of Theorem 2**

Since the violation probability mapping $V(\{ \cdot \})$ is not necessarily upper semi-continuous, the quantity $V(\mathcal{X}_M(\omega)) = \sup_{x \in \mathcal{X}_M(\omega)} V(\{ x \})$ in (3) may not be attained on the compact set $\mathcal{X}_M(\omega)$ [21, Section 1.C]. Therefore we proceed as follows. For all $\omega \in \Delta^N$, from the definition of supremum $\mathcal{V}(\mathcal{X}) = \sup_{x \in \mathcal{X}_M(\omega)} V(\{ x \})$ it holds that for all $\epsilon' > 0$ there exists $\xi^*_M(\omega) \in \mathcal{X}_M(\omega) = \text{conv}(\{ x^1_1(\omega), x^2_1(\omega), ..., x^M_1(\omega) \})$ such that

$$V(\mathcal{X}_M(\omega)) = \sup_{x \in \mathcal{X}_M(\omega)} V(\{ x \}) < V(\xi^*_M(\omega)) + \epsilon'.$$  

### Table 1

<table>
<thead>
<tr>
<th>$n = 3, m = 2, p = 1$</th>
<th>$n = 4, m = 2, p = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>$\xi_{\text{VC}}$ (7)</td>
</tr>
<tr>
<td>0.1</td>
<td>35946</td>
</tr>
<tr>
<td>0.01</td>
<td>530865</td>
</tr>
</tbody>
</table>
Now, for all $\omega \in \Delta^N$, we denote by $I(\omega) \subset \mathbb{Z}[1, M]$ the set of indices of cardinality $|I(\omega)| = \min\{n+1, M\}$, with "minimum lexicographic order", such that we have the inclusion $\xi_M^\omega(\omega) \in \text{conv}\{(x_j^\omega) \mid j \in I(\omega)\}$. Since $X_M(\omega)$ is convex and compact, it follows from Carathéodory's Theorem [29, Theorem 17.1] that such a set of indices $I(\omega)$ always exists. It also follows that there exists a unique set of coefficients $\alpha_1(\omega), \alpha_2(\omega), ..., \alpha_{n+1}(\omega) \in [0, 1]$ such that
\[ \sum_{j \in I(\omega)} \alpha_j(x_j^\omega) = 1 \]
and
\[ \xi_M^\omega(\omega) = \sum_{j \in I(\omega)} \alpha_j(x_j^\omega). \]  
Equation 20

In the following inequalities, we exploit (19), (20), and the convexity of the mapping $x \mapsto g(x, \delta)$ for each fixed $\delta \in \Delta$ from Standing Assumption 1.
\[ \mathbb{P}^N[V(X_M(\omega)) > \epsilon] = \mathbb{P}^N[\sup_{x \in X_M(\omega)} V(x) > \epsilon] \leq \mathbb{P}^N[V(\xi_M^\omega(\omega)) > \epsilon - \epsilon'] \leq \mathbb{P}^N[\sum_{j \in I(\omega)} \alpha_j(x_j^\omega) g(x_j^\omega, \delta) > 0 > \epsilon - \epsilon'] \leq \mathbb{P}^N[\sum_{j \in I(\omega)} \alpha_j(x_j^\omega, \delta) > 0 > \epsilon - \epsilon'] \leq \mathbb{P}^N[N \max_{j \in I(\omega)} \left\{ g(x_j^\omega, \delta) > 0 \right\} > \epsilon - \epsilon'] \leq \mathbb{P}^N[\sum_{j \in I(\omega)} \{ g(x_j^\omega, \delta) > 0 \} > \epsilon - \epsilon'] \leq \mathbb{P}^N[V(\xi_j^\omega(\omega)| j \in I(\omega))] > \epsilon - \epsilon'] \leq \mathbb{P}^N[V(x_j^\omega(\omega)| j \in I(\omega))] > \epsilon - \epsilon'] . \]  
Equation 21

Since for all $k \in \mathbb{Z}[1, M]$, $x_k^\star(\cdot)$ is the optimizer mapping of $\mathbb{S}_P^\prime$ in (8), from [10, Theorem 11], [11, Theorem 3.3] we have that $\mathbb{P}^N(\omega \in \Delta^N | V(x_k^\star(\cdot)) > \epsilon) \leq \Phi(\epsilon, n, N)$. It now follows from Lemma 1 that
\[ \mathbb{P}^N[V(\{x_k^\star(\omega), x_k^\star(\omega), ..., x_M^\star(\omega)\}) > \epsilon'] \leq \frac{M}{n+1} \Phi(\epsilon', n, N). \]

Then, since for all $n, N \geq 1$ the mapping $\epsilon \mapsto \Phi(\epsilon, n, N)$ is continuous, we have that\[ \lim_{\epsilon' \to 0} M \Phi(\epsilon', n, N) = M \Phi(\epsilon, n, N), \]
which proves $\mathbb{P}^N[V(X_M(\omega)) > \epsilon] \leq M \Phi(\epsilon, n, N)$ and in turn (11).

REFERENCES


3With “minimum lexicographic order” we mean the following ordering: \{i_1, i_2, ..., i_n\} < \{j_1, j_2, ..., j_n\} if there exists $k \in \mathbb{Z}[1, n]$ such that $i_1 = j_1, ..., i_{k-1} = j_{k-1}$, and $i_k < j_k$. 