A scenario approach for non-convex control design

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

Abstract

Randomized optimization is an 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 to non-convex problems, but they usually are conservative in terms of performance and require high sample complexity to achieve the desired probabilistic guarantees. In this paper, we derive a novel scenario approach for a wide class of random non-convex programs, with a sample complexity similar to that of uncertain convex programs and with probabilistic guarantees that hold not only for the optimal solution of the scenario program, but for all feasible solutions inside a set of a-priori chosen complexity. We also address measure-theoretic issues for uncertain convex and non-convex programs. Among the family of non-convex control-design problems that can be addressed via randomization, we apply our scenario approach to randomized Model Predictive Control for chance-constrained nonlinear control-affine systems.

I. INTRODUCTION

Modern control design often relies on the solution of an optimization problem, for instance in robust control synthesis [1], [2], Lyapunov-based optimal control [3], [4], and Model Predictive Control (MPC) [5], [6]. 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

The authors are with the Automatic Control Laboratory, ETH Zurich, Switzerland. E-mail addresses: {grammatico, xiaozhan, margellos, pgoulart, jlygeros}@control.ee.ethz.ch.
given specifications for all possible realizations of the uncertainty. See [7] for an example of robust quadratic Lyapunov function synthesis for interval-uncertain linear systems. The worst-case approach is often formulated as a robust optimization problem. However, even though certain classes of robust convex problems are known to be computationally tractable [8], robust convex programs are in general difficult to solve [9], [10]. Moreover, from an engineering perspective, robust solutions generally tend to be conservative in terms of performance.

To reduce the conservativism of robust solutions, stochastic programming [11], [12] offers an alternative methodology. Unlike the worst-case approach, the constraints of the problem can be treated in a probabilistic sense via chance constraints [13], [14], 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 [15], [16]. Unfortunately, even for uncertain convex programs, the feasible set of a chance constraint is in general non-convex, which makes optimization under chance constraints problematic [16, Section 1, pag. 970].

An established and computationally-tractable approach to approximate chance constrained problems is the scenario approximation [15]. A feasible solution of 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, since it does not require any further knowledge on the underlying probability distribution. 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 (so called random) convex programs was first introduced in [17], and many control-design applications are outlined in [18]. The fundamental contribution in these works is the characterization of the number of scenarios, i.e. the sample complexity, needed to guarantee that, with high confidence, the optimal solution of the SP is a feasible solution to the original CCP. The sample complexity bound was further refined in [19] where it was shown to be tight for the class of fully-supported problems, and in [20], [21] where the concept of Helly’s dimension [20, Definition 3.1] and support dimension are
introduced, respectively, to reduce the conservatism for non-fully-supported problems. In [20], [22], the possibility of removing sampled constraints (sampling and discarding) is considered to improve the cost function at the price of decreased feasibility. It is indeed shown that if the constraints of the SP are removed optimally, then the solution of the SP approaches the actual optimal solution of the original CCP.

While feasibility, optimality and sample complexity of random convex programs are well characterized, to the best of the authors’ knowledge, scenario approaches for random non-convex programs are less developed. One family of methods comes from statistical learning theory, based on the Vapnik-Chervonenkis (VC) theory [23], [24], [25], and it is applicable to many non-convex control-design problems [26], [27], [28]. Contrary to scenario approximations of uncertain convex program [19], [20], the aforementioned methods provide probabilistic guarantees for all feasible solutions of the sampled program and not just for the optimal solution. This feature is fundamental because the global optimizer of non-convex programs is not numerically computable in general. Moreover, having probabilistic guarantees for all feasible solutions is of interest in many applications, for instance in [29]. However, the more general probabilistic guarantees of VC theory come at the price of an increased number of required random samples [18]. More fundamentally, they depend on the so-called VC-dimension which is in general difficult to compute, or even infinite, in which case VC theory is not applicable [17].

The aim of this paper is to propose a scenario approach for a wide class of random non-convex programs, with moderate sample complexity, providing probabilistic guarantees for all feasible solutions in a set of a-priori chosen complexity. In the spirit of [17], [18], [19], [20], our results are only based on the “decision complexity”, while no assumption is made on the underlying probability structure. The main contributions of this paper with respect to the existing literature are summarized next.

- We formulate a scenario approach for the class of random non-convex programs with (possibly) non-convex cost, deterministic (possibly) non-convex constraints, and chance constraints containing functions with separable non-convexity. For this class of programs, we show via a counterexample that the standard scenario approach is not directly applicable, because Helly’s dimension (associated with the global optimum) can be unbounded. This motivates the development of our technique.
- We provide a sample complexity bound similar to the one of random convex programs for
all feasible solutions in a set of a-priori chosen degree of complexity.

- We apply our scenario approach methodology to random non-convex programs in the presence of mixed-integer decision variables, with graceful degradation of the associated sample complexity.

- We apply our scenario approach to randomized Model Predictive Control for nonlinear control-affine systems with chance constraints.

- We address the measure-theoretic issues regarding the measurability of the optimal value and optimal solutions of random (convex and non-convex) programs, including the well-definiteness of the probability integrals, under minimal measurability assumptions.

The paper is structured as follows. Section II presents the technical background and the problem statement. Section III presents the main results. Discussions and comparisons with existing methodologies are given in Section IV. Section V presents a scenario approach for randomized MPC of nonlinear control-affine systems. We conclude the paper in Section VI. For ease of reading, the Appendices contain: the analytical example with unbounded Helly’s dimension (Appendix A), the technical proofs (Appendix B), and the measure-theoretic results (Appendix C).

Notation

\( \mathbb{R} \) and \( \mathbb{Z} \) denote, respectively, the set of real and integer numbers. The notation \( \mathbb{Z}[a,b] \) denotes the integer interval \( \{a,a+1,...,b\} \subseteq \mathbb{Z} \). The notation \( \text{conv}(\cdot) \) denotes the convex hull.

II. TECHNICAL BACKGROUND AND PROBLEM STATEMENT

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 \( \mathcal{X} \subset \mathbb{R}^n \):

\[
\text{CCP}(\epsilon) : \begin{cases} 
\min_{x \in \mathcal{X}} J(x) \\
\text{sub. to: } \mathbb{P}(\{\delta \in \Delta \mid g(x,\delta) \leq 0\}) \geq 1 - \epsilon.
\end{cases}
\]

In (1), \( x \in \mathcal{X} \) is the decision variable and \( \delta \) is a random variable defined on a probability space \( (\Delta, \mathcal{F}, \mathbb{P}) \), with \( \Delta \subseteq \mathbb{R}^m \). All the measure-theoretic arguments associated with the probability measure \( \mathbb{P} \) are addressed in Appendix C.

Throughout the paper, we make the following assumption, partially adopted from [20, Assumption 1].
**Standing Assumption 1** (Regularity). The set \( \mathcal{X} \subseteq \mathbb{R}^n \) is compact and convex. For all \( \delta \in \Delta \subseteq \mathbb{R}^m \), the mapping \( x \mapsto g(x, \delta) \) is convex and lower semicontinuous. For all \( x \in \mathbb{R}^n \), the mapping \( \delta \mapsto g(x, \delta) \) is measurable. The function \( J \) is lower semicontinuous.

The compactness assumption on \( \mathcal{X} \), typical of any problem of practical interest, avoids technical difficulties by guaranteeing that any feasible problem instance attains a solution \([20, Section 3.1, pag. 3433]\). The set \( \mathcal{X} \) is assumed convex without loss of generality\[1\]. Measurability of \( g(x, \cdot) \) and lower semicontinuity of \( J \) are needed to avoid potential measure-theoretic issues, see Appendix \[C\] for technical details.

**Remark 1.** The CCP formulation in (1) implicitly includes the more general CCP

\[
\text{CCP}'(\epsilon) : \begin{cases} 
\min_{x \in \mathcal{X}} J(x) \\
\text{sub. to: } \mathbb{P} \left( \{ \delta \in \Delta \mid g(x, \delta) + f(x)\varphi(\delta) \leq 0 \} \right) \geq 1 - \epsilon \\
h(x) \leq 0,
\end{cases}
\]

for possibly non-convex functions \( f, h : \mathbb{R}^n \to \mathbb{R}, \varphi : \mathbb{R}^m \to \mathbb{R} \), at the only price of introducing one extra variable\[2\] \( y = f(x) \in \mathcal{Y} := f(\mathcal{X}) \).

It is important to notice that unlike the standard setting of random convex programs \([17]\), we allow the cost function \( J \) to be non-convex. Since our results presented later on provide probabilistic guarantees for an entire set, rather than for a single point, we next give the set-based counterpart of \([18, Definitions 1, 2]\).

**Definition 1** (Probability of Violation and Feasibility of a Set). The **probability of violation** of a set \( \mathcal{X} \subseteq \mathcal{X}' \) is defined as

\[
V(\mathcal{X}) := \sup_{x \in \mathcal{X}} \mathbb{P} \left( \{ \delta \in \Delta \mid g(x, \delta) > 0 \} \right). 
\]

\[1\] If \( \mathcal{X} \) is not convex, let \( \mathcal{X}' \supset \mathcal{X} \) be a compact convex superset of \( \mathcal{X} \). Then we can define the indicator function \( \chi : \mathbb{R}^n \to \{0, \infty\} \), see \([30\ Section 1.A, pag. 6–7]\) as \( \chi(x) := 0 \) if \( x \in \mathcal{X}' \), \( \infty \) otherwise. Then we define the new cost function \( J + \chi \), which is lower semicontinuous as well, and finally consider the CCP with convex feasibility set \( \mathcal{X}' \) as \( \min_{x \in \mathcal{X}'} J(x) + \chi(x) \) sub. to: \( \mathbb{P} \left( \{ \delta \in \Delta \mid g(x, \delta) \leq 0 \} \right) \geq 1 - \epsilon \).

\[2\] We follow the lines of \([30\ Section 1.A, pag. 6–7]\). The probabilistic constraint becomes \( \mathbb{P} \left( \{ \delta \in \Delta \mid g(x, \delta) + y\varphi(\delta) \leq 0 \} \right) \geq 1 - \epsilon \), while the deterministic constraint becomes \( \max \{h(x), |y - f(x)|\} \leq 0 \). We now define the indicator function \( \chi : \mathcal{X} \times \mathbb{R} \to \{0, \infty\} \) as \( \chi(x, y) := 0 \) if \( \max \{h(x), |y - f(x)|\} \leq 0 \), \( \infty \) otherwise, in order to get a CCP formulation as in (1). More generally, we allow for “separable” probabilistic constraint of the kind \( \mathbb{P} \left( \{ \delta \in \Delta \mid \phi(g(x, \delta) + \sum_i f_i(x)\varphi(\delta)) \leq 0 \} \right) \geq 1 - \epsilon \), for convex \( \phi : \mathbb{R}^{p \times q} \to \mathbb{R} \), and possibly non-convex functions \( f_i : \mathbb{R}^n \to \mathbb{R} \).
For any given $\epsilon \in (0, 1)$, a set $X \subseteq \mathcal{X}$ is feasible for CCP($\epsilon$) in (1) if $V(X) \leq \epsilon$. 

In view of Definition 1, which accounts for the worst-case violation probability on an entire set, our developments are partially inspired by the following key statement regarding the violation probability of the convex hull of a given set.

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

To the best of our knowledge this basic fact has not been exploited in the literature. An immediate consequence of Theorem 1 is that the feasibility set $X_{\epsilon} := \{x \in \mathcal{X} \mid P(\{\delta \in \Delta \mid g(x, \delta) \leq 0\}) \geq 1 - \epsilon\}$ of CCP($\epsilon$) in (1) satisfies

$$X_{\epsilon} \subseteq \text{conv}(X_{\epsilon}) \subseteq X_{(n+1)\epsilon}.$$ 

Associated with CCP($\epsilon$) in (1), we consider a Scenario Program (SP) obtained from $N$ independent and identically distributed (i.i.d.) samples $\bar{\delta}(1), \bar{\delta}(2), \ldots, \bar{\delta}(N)$ drawn according to $\mathbb{P}$ [17, Definition 3]. For a fixed multi-sample $\bar{\omega} := (\bar{\delta}(1), \bar{\delta}(2), \ldots, \bar{\delta}(N)) \in \Delta^N$, we consider the SP

$$\text{SP}[\bar{\omega}] : \begin{cases} \min_{x \in \mathcal{X}} J(x) \\ \text{sub. to: } g(x, \bar{\delta}(i)) \leq 0 \quad \forall i \in \mathbb{Z}[1,N]. \end{cases}$$

(4)

A. Known results on scenario approximations of chance constraints

In [19], [20], the case $J(x) := c^\top x$ is considered. Under the assumption that, for every multi-sample, the optimizer is unique [19, Assumption 1], [20, Assumption 2] or a suitable tie-breaking rule is adopted [17, Section 4.1] [19, Section 2.1], the optimizer mapping $x^*(\cdot) : \Delta^N \rightarrow \mathcal{X}$ of SP[·] is such that

$$\mathbb{P}^N\left(\{\omega \in \Delta^N \mid V(\{x^*(\omega)\}) > \epsilon\}\right) \leq \Phi(\epsilon, n, N) := \sum_{j=0}^{n-1} \binom{N}{j} \epsilon^j (1 - \epsilon)^{N-j}. \quad (5)$$

The above bound is tight for fully-supported problems [19, Theorem 1, Equation (7)], while for non-fully-supported problems it can be improved by replacing $n$ with the so-called Helly’s dimension $\zeta \leq n$ [20, Theorem 3.3]. To satisfy the implicit bound (5) with right-hand side equal to $\beta \in (0, 1)$, it is sufficient to select a sample size [20, Corollary 5.1], [31, Lemma 2]

$$N \geq \frac{e-1}{\epsilon} \left( \zeta - 1 + \ln \left( \frac{1}{\beta} \right) \right). \quad (6)$$
We emphasize 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 (i.e. with $J(x) := c^\top x$) is the possibly non-convex cost $J$, we show in Appendix A that Helly’s dimension $\zeta$ for the globally optimal value of SP in (4) can in general be unbounded. Therefore even for the apparently simple non-convex SP in (4) it is impossible to directly apply the classic scenario approach [17], [18], [19], [20] based on Helly’s theorem [32], [33].

For general non-convex programs, VC theory provides upper bounds for the quantity $P_N(\{\omega \in \Delta^N | V(X(\omega)) > \epsilon\})$, where $X(\omega) \subseteq \mathcal{X}$ is the entire feasible set of SP[\omega], see the discussions in [34, Section 3.2], [26, Sections IV, V].

In particular, [24, Theorem 8.4.1] shows that it suffices to select a sample size

$$N_{VC} \geq \frac{4}{\epsilon} \left( \xi \ln \left( \frac{12}{\epsilon} \right) + \ln \left( \frac{2}{\beta} \right) \right),$$

(7)

to guarantee with confidence $1 - \beta$ that any feasible solution of SP[\omega] has probability of violation no larger than $\epsilon$. In (7), $\xi$ is the so-called VC dimension [28, Definition 10.2], which encodes the richness of the family of functions $\{\delta \mapsto g(x, \delta) \mid x \in \mathcal{X}\}$ and may be hard to estimate, or even infinite.

III. RANDOM NON-CONVEX PROGRAMS: PROBABILISTIC GUARANTEES FOR AN ENTIRE SET

A. Main results

We start with a preliminary intuitive statement. We consider a finite number of mappings $x^*_1, x^*_2, \ldots, x^*_M : \Delta^N \to \mathcal{X}$, each one with probabilistic guarantees, and we upper bound their worst-case probability of violation.

**Assumption 1.** For given $\epsilon \in (0, 1)$, the mappings $x^*_1, x^*_2, \ldots, x^*_M : \Delta^N \to \mathcal{X}$ are such that, for all $k \in \mathbb{Z}[1, M]$, we have $P_N(\{\omega \in \Delta^N | V(\{x^*_k(\omega)\}) > \epsilon\}) \leq \beta_k \in (0, 1)$.

**Lemma 1.** Consider the SP[\omega] in (4) with $N \geq n$. If Assumption [1] holds, then

$$P_N (\{\omega \in \Delta^N | V(\{x^*_1(\omega), x^*_2(\omega), \ldots, x^*_M(\omega)\}) > \epsilon\}) \leq \sum_{k=1}^{M} \beta_k.$$
the set \( \{ x_1^*(\omega), x_2^*(\omega), \ldots, x_M^*(\omega) \} \) is feasible with respect to Definition 1, i.e.,

\[
P_N \left( \{ \omega \in \Delta^N \mid V \left( \{ x_1^*(\omega), x_2^*(\omega), \ldots, x_M^*(\omega) \} \right) \leq \epsilon \} \right) \geq 1 - M\beta.
\]

We may consider the meaning of Lemma 1 in view of the result in [24, Section 4.2], and similarly in [31, Section 4.2], where the decision variable \( x \) lives in a set \( \mathcal{X} \) of finite cardinality. The main difference here is that Lemma 1 instead relies on a finite number of mappings \( x_k^*(\cdot) \), rather than on a finite number of decisions. Each of these mappings is associated with a given upper bound \( \beta_k \) on the probability of violating the chance constraint.

We now proceed to our main idea. We address the CCP(\( \epsilon \)) in (1) through a family of \( M \) distinct convex SPs, each with Helly’s dimension bounded by some integer \( \zeta \in \mathbb{Z}[1, n] \). We consider \( M \) cost vectors \( c_1, c_2, \ldots, c_M \in \mathbb{R}^n \), and for each \( k \in \mathbb{Z}[1, M] \), we define the following SP, where \( \bar{\omega} := (\bar{\delta}(1), \bar{\delta}(2), \ldots, \bar{\delta}(N)) \).

\[
\text{SP}_k[\bar{\omega}] : \begin{align*}
\min_{x \in \mathcal{C}_k \cap \mathcal{X}} & \quad c_k^T x \\
\text{sub. to:} & \quad g(x, \delta(i)) \leq 0 \quad \forall i \in \mathbb{Z}[1, N]
\end{align*}
\]  (8)

The additional convex constraint \( x \in \mathcal{C}_k \subseteq \mathbb{R}^n \) allows to upper bound Helly’s dimension by some \( \zeta \in \mathbb{Z}[1, n] \), and its choice is hence discussed later on. Let \( x_k^*(\bar{\omega}) \) be the optimizer of \( \text{SP}_k[\bar{\omega}] \) and assume that it is unique, or a suitable tie-break rule is considered [17, Section 4.1]. We adopt the convention that \( x_k^*(\bar{\omega}) := \emptyset \) if \( \text{SP}_k[\bar{\omega}] \) is not feasible. We notice that if \( \text{SP}_k[\bar{\omega}] \) is feasible then the optimizer \( x_k^*(\bar{\omega}) \) is always finite due to the compactness assumption on \( \mathcal{X} \).

For all \( \omega \in \Delta^N \), let us consider the convex-hull set

\[
\mathbb{X}_M(\omega) := \text{conv} \left( \{ x_1^*(\omega), x_2^*(\omega), \ldots, x_M^*(\omega) \} \right),
\]  (9)

where, for all \( k \in \mathbb{Z}[1, M] \), \( x_k^*(\cdot) \) is the optimizer mapping of \( \text{SP}_k[\cdot] \) in (8). The construction of \( \mathbb{X}_M \) is illustrated in Figure 1.

We are now ready to state our main result.

**Theorem 2.** For each \( k \in \mathbb{Z}[1, M] \), let \( x_k^* \) and \( \zeta \) be, respectively, the optimizer mapping and an upper bound for Helly’s dimension of \( \text{SP}_k \) in (8), and let \( \mathbb{X}_M \) be as in (9). Then, for all \( \epsilon \in (0, 1) \),

\[
P_N \left( \{ \omega \in \Delta^N \mid V(\mathbb{X}_M(\omega)) > \epsilon \} \right) \leq M\Phi \left( \frac{\epsilon}{\min\{n + 1, M\}}, \zeta, N \right). \]  (10)
Fig. 1. The set $\mathbb{X}_M$ is the convex hull of the points $x_1, x_2, ..., x_M$, where each $x_k$ is the optimizer of $\text{SP}_k$ in (8) having linear cost $c_k^\top x$.

Following the lines of [35, Appendix, Proof of Theorem 2], we can also slightly improve the bound of Theorem 2 as follows.

**Corollary 1.** For each $k \in \mathbb{Z}[1, M]$, let $x_k^*$ and $\zeta$ be, respectively, the optimizer mapping and Helly’s dimension of $\text{SP}_k$ in (8), and let $\mathbb{X}_M$ be as in (9). Then, for all $\epsilon \in (0, 1)$,

$$
\mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(\mathbb{X}_M(\omega)) > \epsilon \} \right) \leq \left( \frac{M}{n + 1} \right) \Phi \left( \epsilon, \zeta \min \{ n + 1, M \}, N \right).
$$

(11)

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

$$
\tilde{\text{SP}}[\bar{\omega}]: \begin{cases} 
\min_{x \in \mathcal{X}} J(x) \\
\text{sub. to: } x \in \mathbb{X}_M(\bar{\omega}),
\end{cases}
$$

(12)

and explicitly compute the corresponding sample complexity.
Corollary 2. For each \( k \in \mathbb{Z}[1, M] \), let \( x^*_k \) and \( \zeta \) be, respectively, the optimizer mapping and Helly’s dimension of \( \text{SP}_k \) in (8), and let \( X_M \) be as in (9). For all \( \epsilon, \beta \in (0, 1) \), if
\[
N \geq \frac{e}{\epsilon} \min\{n + 1, M\} \left( \zeta - 1 + \ln \left( \frac{M}{\beta} \right) \right),
\]
then \( \mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(X_M(\omega)) \leq \epsilon \} \right) \geq 1 - \beta \), i.e., with probability no smaller than \( 1 - \beta \), any feasible solution of \( \text{SP}[\bar{\omega}] \) in (12) is feasible for \( \text{CCP}(\epsilon) \) in (1).

Remark 2. The constraint \( x \in C_k \) in (8) provides a way to upper bound Helly’s dimension \( \zeta \) of \( \text{SP}_k[\bar{\omega}] \). Many choices of \( C_k \) are possible. For instance, \( C_k := \mathbb{R}^n \) in general only provides the upper bound \( \zeta \leq n \). The minimum upper bound on Helly’s dimension for \( \text{SP}_k[\bar{\omega}] \) in (8), i.e. \( \zeta = 1 \), is obtained whenever \( x \) is constrained to live in a linear subspace of dimension one [21, Lemma 3.8]. This happens independently from \( g \) if, for some fixed \( x^0, c_k \in \mathbb{R}^n \), \( k = 1, 2, ..., M \), we define
\[
C_k := \{ x^0_k + \lambda c_k \in \mathbb{R}^n \mid \lambda \in \mathbb{R} \}.
\]

With such a choice of \( C_k \), \( \text{SP}_k \) is equivalent to the program
\[
\min_{\lambda \in \mathbb{R}} \{ -\lambda \} \quad \text{sub. to:} \quad (x^0 + \lambda c_k) \in \mathcal{X}, \quad g(x^0 + \lambda c_k, \delta^i) \leq 0 \quad \forall i \in \mathbb{Z}[1, N],
\]
which has unique optimizer and Helly’s dimension \( \zeta = 1 \), since the decision variable \( \lambda \) is 1-dimensional. In this case, the required sample size (for \( M \geq n + 1 \)) from (13) is \( (\frac{e}{e-1}) (n + 1)/\epsilon \ln (M/\beta) \), which is linear in the number \( n \) of decision variables. In particular, if we a-priori know a feasible point \( x^0 \) for \( \text{CCP}(\epsilon) \) in (1), then the solution of (15) generates a point \( x^0 + \lambda_k^*(\bar{\omega}) c_k \in X(\bar{\omega}) \) for each \( k \in \mathbb{Z}[1, M] \). This additional knowledge is available in many situations of interest [36, Section 1.1], for instance in [37], [38].

We note that other convex problems can be used in place of \( \text{SP}_k \) in (8). For instance, consider the set \( \mathbb{R}B := \{ x \in \mathbb{R}^n \mid \|x\| \leq R \} \subseteq \mathbb{R}^n \), where \( R > 0 \) is such that \( \mathcal{X} \subset \mathbb{R}B \), and \( M \) points \( z_1, z_2, ..., z_M \) on the boundary of \( \mathbb{R}B \). For each \( k \in \mathbb{Z}[1, M] \), we can define the following SP.
\[
\text{SP}'_k[\bar{\omega}] := \left\{ \min_{x \in \mathcal{X}} \|x - z_k\| \quad \text{sub. to:} \quad g(x, \delta^{(i)}) \leq 0 \quad \forall i \in \mathbb{Z}[1, N]. \right\}
\]
More generally, the way of selecting the convex problems \( \text{SP}_k[\bar{\omega}] \), and hence their associated optimizers \( x^*_k(\bar{\omega}) \), for \( k = 1, 2, ..., M \), is not an essential feature for our feasibility results. □
In view of the approximating, non-convex, problem $\tilde{\text{SP}}[\bar{\omega}]$ in (12) we are interested in getting a “large” $X_M(\bar{\omega})$ in (9). The choice in (8) is motivated by the fact that the optimal solution $x_k^*(\bar{\omega})$ of $\text{SP}_k[\bar{\omega}]$ belongs to the boundary of the actual (convex) feasibility set $X(\bar{\omega}) := \{x \in \mathcal{X} \mid g(x, \bar{\delta}^{(i)}) \leq 0 \ \forall i \in \mathbb{Z}[1, N]\}$, therefore so do the extreme points of the convex-hull set $X_M$ in (9), as shown in Figure 1.

We finally emphasize that we obtain the probabilistic guarantees in (2) for any feasible solution of $\tilde{\text{SP}}$ in (12), not just for the optimal solution. This is of practical importance, because $\tilde{\text{SP}}[\bar{\omega}]$ is non-convex and hence it is in general impossible to numerically compute its optimal solution.

\section*{B. On mixed-integer random non-convex programs}

The results in Lemma 1 and Theorem 2 can be further exploited to provide probabilistic guarantees for the following class of mixed-integer CCPs.

\[
\text{CCP}_{\text{m-i}}(\epsilon) : \begin{cases} 
\min_{(x,j) \in \mathcal{X} \times \mathbb{Z}[1,L]} & J(x) \\
\text{sub. to:} & P\left(\{\delta \in \Delta \mid g_j(x, \delta) \leq 0\}\right) \geq 1 - \epsilon,
\end{cases}
\]

(17)

where the functions $g_1, g_2, \ldots, g_L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ satisfy the following assumption.

\textbf{Assumption 2.} For all $j \in \mathbb{Z}[1,L]$ and $\delta \in \Delta$, the mapping $x \mapsto g_j(x, \delta)$ is convex and lower semicontinuous.

Notice that unlike [39], [40], we also allow for possibly non-convex objective functions $J$.

We also define the probability of violation (of any set $X \subseteq \mathcal{X}$) associated with $\text{CCP}_{\text{m-i}}(\epsilon)$ in (17) as

\[
V_{\text{m-i}}(X) := \sup_{x \in \mathcal{X}} P\left(\{\delta \in \Delta \mid \min_{j \in \mathbb{Z}[1,L]} g_j(x, \delta) > 0\}\right).
\]

(18)

Note that, for all $j \in \mathbb{Z}[1,L]$, it holds $V_{\text{m-i}}(X) \leq \sup_{x \in \mathcal{X}} P\left(\{\delta \in \Delta \mid g_j(x, \delta) > 0\}\right)$.

We can proceed similarly to Section III-A. For fixed multi-sample $\bar{\omega} \in \Delta^N$, we consider the $M$ cost vectors $c_1, c_2, \ldots, c_M \in \mathbb{R}^n$ and the convex sets $C_1, C_2, \ldots, C_M \subseteq \mathbb{R}^n$, so that, for all $(j,k) \in \mathbb{Z}[1,L] \times \mathbb{Z}[1,M]$ we define

\[
\text{SP}_{j,k}^{\text{m-i}}[\bar{\omega}] : \begin{cases} 
\min_{x \in C_k \cap \mathcal{X}} c_k^T x \\
\text{sub. to:} & g_j(x, \bar{\delta}^{(i)}) \leq 0 \ \forall i \in \mathbb{Z}[1,N]
\end{cases}
\]

(19)

with optimizer $x_{j,k}^*(\bar{\omega})$. Then we can define the set $X_j(\omega)$ as in (8)–(9), i.e.

\[
X_j(\omega) := \text{conv}\left(\{x_{j,1}^*(\omega), x_{j,2}^*(\omega), \ldots, x_{j,M}^*(\omega)\}\right).
\]

(20)
If $\zeta \in \mathbb{Z}[1, n]$ is an upper bound for Helly’s dimension of the convex programs $SP_{j,k}^{m-i}[\omega]$, then it follows from Theorem 2 and (18) that, for all $j \in \mathbb{Z}[1, L]$, we have

$$
P^N \left( \{ \omega \in \Delta^N \mid V^{m-i}(X_j(\omega)) > \epsilon \} \right) \leq \mathbb{P}^N \left( \{ \omega \in \Delta \mid \sup_{x \in X_j(\omega)} \mathbb{P} \left( \{ \delta \in \Delta \mid g_j(x, \delta) \right) > \epsilon \} \right) \leq M\Phi \left( \frac{\epsilon}{\min\{n+1, M\}}, \zeta, N \right). \quad (21)$$

We can then establish the following upper bound on the probability of violation of the union of convex-hull sets constructed above.

**Theorem 3.** Suppose Assumption 2 holds. For each $(j, k) \in \mathbb{Z}[1, L] \times \mathbb{Z}[1, M]$, let $x^*_{j,k}$ and $\zeta$ be, respectively, the optimizer mapping and an upper bound for Helly’s dimension of $SP_{j,k}^{m-i}$ in (19); let $X_j$ be defined as in (20). Then

$$
P^N \left( \{ \omega \in \Delta^N \mid V^{m-i}(\bigcup_{j=1}^L X_j(\omega)) > \epsilon \} \right) \leq LM\Phi \left( \frac{\epsilon}{\min\{n+1, M\}}, \zeta, N \right). \quad (22)$$

We can now approximate the CCP$^{m-i}(\epsilon)$ in (17) by

$$
\tilde{SP}^{m-i}[\omega] : \left\{ \begin{array}{l}
\min_{(x,j) \in X \times \mathbb{Z}[1, L]} J(x) \\
\text{sub. to: } x \in X_j(\omega),
\end{array} \right. \quad (23)
$$

and state the following lower bound on the required sample size.

**Corollary 3.** Suppose Assumption 2 holds. For each $(j, k) \in \mathbb{Z}[1, L] \times \mathbb{Z}[1, M]$, let $x^*_{j,k}$ and $\zeta$ be, respectively, the optimizer mapping and an upper bound for Helly’s dimension of $SP_{j,k}^{m-i}$ in (19); let $X_j$ be defined as in (20). If

$$
N \geq \frac{\epsilon}{\epsilon - 1} \min\{n + 1, M\} \left( \zeta - 1 + \ln \left( \frac{LM}{\beta} \right) \right), \quad (24)
$$

then $\mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V^{m-i}(\bigcup_{j=1}^L X_j(\omega)) > \epsilon \} \right) \geq 1 - \beta$, i.e., with probability no smaller than $1 - \beta$, any feasible solution of $\tilde{SP}^{m-i}[\omega]$ in (23) is feasible for CCP$^{m-i}(\epsilon)$ in (17). \hfill \Box

Let us comment on the sample size $N$ given in Corollary 3 relative to SP in (23). The formulation in (17) subsumes the ones in [39] and [40]. In [40] Section 4 it is shown that it is possible to derive a sample size $N$ which grows linearly with the dimension $d$ of the integer variable $y \in (\mathbb{Z}[-l/2, l/2])^d$, so that $L := (l + 1)^d$ in (17). The addition here is that we can also
deal with non-convex objective functions \( J(x) \) and non-convex deterministic constraints \( h(x) \leq 0 \) according to Remark 1, still maintaining a sample size with logarithmic dependence on \( L \), i.e. linear dependence on \( d \). The result in [39, Theorem 3] presents an exponential dependence of the sample size \( N \) with respect to the dimension \( d \) of the integer variable \( y \), but the result therein is slightly more general because it technically covers for a possibly unbounded domain for \( y \).

IV. DISCUSSION AND COMPARISONS

A. Sampling and discarding

The problem \( \text{SP}_k \) in (8) is also suitable for a sampling-and-discarding approach [20], [22]. In particular, the aim is to reduce the optimal value of each (convex) \( \text{SP}_k \) in (8), and hence enlarge the set \( \mathbb{X}_M \) in (9). Indeed, we can a-priori decide that we will discard \( r \) of the \( N \) samples of the uncertainty. As discussed in [20], any removal algorithm could be employed for the discarding part. Since the optimal constraint discarding is of combinatorial complexity, [20, Section 5.1] proposes greedy algorithms and an approach based on the Lagrange multipliers associated with the constraint functions. If \( N \) and \( r \) are taken such that

\[
\left( \begin{array}{c}
\zeta + r - 1 \\
r
\end{array} \right) \Phi \left( \epsilon, \zeta + r, N \right) = \left( \begin{array}{c}
\zeta + r - 1 \\
r
\end{array} \right) \sum_{i=1}^{\zeta+r-1} \left( \begin{array}{c}
N \\
i
\end{array} \right) \epsilon^i (1 - \epsilon)^{N-i} \leq \beta,
\]

(25)

where \( \zeta \in \mathbb{Z}[1, n] \) is an upper bound on the Helly’s dimension of the problem \( \text{SP}_k \) in (8), then, for all \( k \in \mathbb{Z}[1, M] \), we have that the optimizer mapping \( x^*_k(\cdot) \) of \( \text{SP}_k[\cdot] \) in (8) (where only \( N-r \) constraints are enforced) is such that \( \mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(\{ x^*_k(\omega) \} > \epsilon) \} \right) \leq \beta \) [20, Theorem 4.1], [22, Theorem 2.1]. Explicit bounds on the sample and removal couple \( (N, r) \) are given in [20, Section 5], [22, Section 4.3].

It then follows from (25) that, with \( r \) removals over \( N \) samples, the optimizer mappings \( x^*_1, x^*_2, \ldots, x^*_M \) satisfy Assumption 1 with \( \beta_k := \left( \begin{array}{c}
\zeta + r - 1 \\
r
\end{array} \right) \Phi \left( \epsilon, \zeta + r, N \right) \) for all \( k \in \mathbb{Z}[1, M] \). Therefore, in view of Lemma 1, we get that the probabilistic guarantees established in Theorem 2 become

\[
\mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(\mathbb{X}_M(\omega)) > \epsilon \} \right) \leq M \left( \begin{array}{c}
\zeta + r - 1 \\
r
\end{array} \right) \Phi \left( \frac{\epsilon}{n+1}, \zeta + r, N \right).
\]

Since the above inequality relies on \( \mathbb{P}^N \), we emphasize that it is possible to remove different sets of \( r \) samples from each \( \text{SP}_k \). Namely, for all \( k \in \mathbb{Z}[1, M] \), let \( \mathcal{I}_k \subseteq \mathbb{Z}[1, M] \) be a set of indices with cardinality \( |\mathcal{I}_k| = r \). Thus, we can discard the samples \( \{ \delta^{(i)} \mid i \in \mathcal{I}_k \} \) from \( \text{SP}_k \), possibly with \( \mathcal{I}_k \neq \mathcal{I}_j \) for \( k \neq j \).
B. Comparison with the statistical learning theory approach

Let us compare our sample size in Corollary 2 to the corresponding bounds from statistical learning theory based on the VC dimension. First, in terms of constraint violation tolerance $\varepsilon$, our sample size in (13) grows as $1/\varepsilon$ while the sample size provided via the classic statistical learning theory grows as $1/\varepsilon^2 \log(1/\varepsilon^2)$ [41, Sections 4, 5], [28, Chapter 8]. An important refinement over the classic result is possible considering the so-called “one-sided probability of constrained failure”,” see for instance [24, Chapter 8], [25, Chapter 7], [34, Section 3], [26, Sections IV, V]. The typical sample size provided in those references is $4/\varepsilon (\xi \log_2 (12/\varepsilon) + \log_2 (2/\beta))$, where $\xi$ is the VC dimension associated with the family of constraint functions $\{g(x, \cdot) : \Delta \to \mathbb{R} | x \in \mathcal{X}\}$. Note that the asymptotic dependence on $\varepsilon$ drops from $1/\varepsilon^2 \log(1/\varepsilon^2)$ to $1/\varepsilon \ln(1/\varepsilon)$, but still remains higher than the sample size in (13). Second, the sampling-and-discarding approach can be used to enlarge the feasibility domain $\mathbb{X}_M(\bar{\omega})$ in (9), as the explicit sample size only grows linearly with the number of removals $r$ [20, Corollary 5.1]. On the other hand [24, Chapter 8, pag. 103], statistical learning theory approaches cover the possibility of discarding a certain fraction $\rho \in [0, 1)$ of the samples, resulting in a sample size of the order of $(\rho + \varepsilon)/\varepsilon^2 \ln((\rho + \varepsilon)/\varepsilon^2)$. Let us indeed denote by $\mathbb{X}_M^r(\bar{\omega})$ the feasibility set of the methodology of Section III-A defined as in (9), but where each vertex $x_k^r(\bar{\omega})$ is computed considering only $N - r$ samples. It then follows that without any discarding, i.e. for $r = 0$, the set $\mathbb{X}_M^0(\bar{\omega}) := \mathbb{X}_M(\bar{\omega})$ in (9) is always a subset of the entire feasibility set $\mathbb{X}(\bar{\omega})$ for any given multi-sample $\bar{\omega} \in \Delta^N$. However, for $r > 0$, the inclusion $\mathbb{X}_M^r(\cdot) \subseteq \mathbb{X}(\cdot)$ is no more true therefore the feasibility set constructed in Section III-A together with a sampling-and-discarding approach, is not necessarily a subset of the classic statistical learning theory counterpart. Third, and most important, the sample size in (13) depends only on the dimension $n$ of the decision variable, not on the VC dimension $\xi$ of the constraint function $g$ and, as already mentioned, $\xi$ may be difficult to estimate, or even infinite, in which case VC theory is not applicable.

On the other hand, approaches based on statistical learning theory offer some advantages over our method. They in fact cover general non-convex problems and, without any sampling and discarding, provide probabilistic guarantees for all feasible points, not only for those in a certain subset of given complexity.
C. Comparison with mixed random-robust approach

An alternative approach based on a mixture of randomized and robust optimization was presented in [42]. It requires solving a robust problem with the uncertainty being confined in an appropriately parametrized set, generated in a randomized way to include \((1 - \epsilon)\) of the probability mass of the uncertainty, with high confidence. Following this approach one obtains probabilistic guarantees for any feasible solution of the robust problem. In particular, the size of this subset depends on the parametrization of the uncertainty set, which in turn affects the number of scenarios that must be extracted. However, in contrast to the current paper, the approach in [42] has some drawbacks listed as follows. First, it is not guaranteed that the a-priori chosen parametrization generates a feasible robust optimization problem. Second, if such robust program is feasible, it is in general conservative in terms of cost and computationally tractable only for a very limited class of non-convex problems. In particular, some additional structure on the dependence on the uncertainty must be assumed. Finally, the method in [42] comes with no explicit characterization of the probabilistically-feasible subset in the decision variable domain.

V. RANDOMIZED MODEL PREDICTIVE CONTROL OF NONLINEAR CONTROL-AFFINE SYSTEMS AND OTHER CONTROL APPLICATIONS

A. Randomized Model Predictive Control of nonlinear control-affine systems

In this section we extend the results of [43], [44] to uncertain nonlinear control-affine systems of the form

\[
x^{+} = f(x, v) + g(x, v)u,
\]

where \(x \in \mathbb{R}^n\) is the state variable, \(u \in \mathbb{R}^m\) is the control variable, and \(v \in \mathcal{V} \subseteq \mathbb{R}^p\) is the uncertain random input. We assume state and control constraints \(x \in \mathcal{X} \subseteq \mathbb{R}^n, u \in \mathcal{U} \subseteq \mathbb{R}^m\), where \(\mathcal{X}\) and \(\mathcal{U}\) are compact convex sets. We further assume the availability of i.i.d. samples \(\bar{v}^{(1)}, \bar{v}^{(2)}, \ldots\) of the uncertain input, drawn according to a possibly-unknown probability measure \(\mathbb{P}\) [17, Definition 3].

For a horizon length \(K\), let \(u := (u_0, u_1, \ldots, u_{K-1})\) and \(v := (v_0, v_1, \ldots, v_{K-1})\) denote a control-input and random-input sequence respectively. We denote by \(\phi(k; x, u, v)\) the state solution of (26) at time \(k \geq 0\), starting from the initial state \(x\), under the control-input sequence \(u\) and the random-input sequence \(v\). Likewise, given a control law \(\kappa : \mathcal{X} \to \mathcal{U}\), we denote by \(\phi_\kappa(k; x, v)\)
the state solution of the system \( x^+ = f(x, v) + g(x, v)\kappa(x) \) at time \( k \geq 0 \), starting from the initial state \( x \), under the random-input sequence \( v \). The solution \( \phi(k; x, u, v) \), as well as \( \phi_n(k; x, v) \), is a random variable itself because under the dependence on the random-input sequence \( v \).

Let \( \ell : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_{\geq 0} \) be the stage cost, and \( \ell_f : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \) be the terminal cost. We consider the random finite-horizon cost function

\[
J(x, u, v) := \ell_f (\phi(K; x, u, v)) + \sum_{k=0}^{K-1} \ell (\phi(k; x, u, v), u_k).
\]

Following [44, Section 3.1], we formulate the multi-stage Stochastic MPC (SMPC) problem

\[
\begin{aligned}
&\min_{u \in \mathbb{U}^K} \mathbb{E}^K [J(x, u, \cdot)] \\
&\text{sub. to: } \mathbb{P}^k \left( \{v \in \mathbb{V}^k \mid \phi(k; x, u, v) \in \mathbb{X} \} \right) \geq 1 - \epsilon \quad \forall k \in \mathbb{Z}[1, K]
\end{aligned}
\]

and its randomized (non-convex) counterpart

\[
\text{SP}^{\text{MPC}}[\bar{v}^{(1)}, \bar{v}^{(2)}, \ldots] : \quad \begin{aligned}
&\min_{u \in \mathbb{U}^K} \sum_{i \in \mathcal{I}_0} J(x, u, \bar{v}^{(i)}) \\
&\text{sub. to: } \phi(1; x, u, \bar{v}^{(i)}) \in \mathbb{X} \quad \forall i \in \mathcal{I}_1 \\
&\phi(k; x, u, \bar{v}^{(i)}) \in \mathbb{X} \quad \forall i \in \mathcal{I}_2, \quad \forall k \in \mathbb{Z}[2, K],
\end{aligned}
\]

for some disjoint index sets \( \mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2 \subset \mathbb{Z}[1, \infty) \). The receding horizon control policy is defined as follows. For each time step, we measure the state \( x \) and let \( u^*(x) := (u_0^*, \ldots, u_{K-1}^*) (x) \) be the solution of \( \text{SP}^{\text{MPC}} \) in (29), for some drawn samples \( \{\bar{v}^{(1)}, \bar{v}^{(2)}, \ldots\} \). The control input \( u \) is set to the first element of the computed sequence, namely \( u = \kappa(x) := u_0^* \), which implicitly also depends on the samples extracted to build the optimization program itself.

We next focus on a suitable choice for the sample size, so that the average fraction of closed-loop constraint violations "\( x_1 \notin \mathbb{X}, x_2 \notin \mathbb{X}, \ldots, x_t \notin \mathbb{X} \)" is below the desired level \( \epsilon \). It follows from [44, Section 3] that this property is actually independent from the cardinalities of \( \mathcal{I}_0 \) and \( \mathcal{I}_2 \), i.e. on the number of samples used for the cost function and for the later stages. In fact, under proper assumptions introduced later on, the closed-loop behavior in terms of constraint violations is only influenced by the first-stage constraint, namely by the number \( N \) of samples indexed in \( \mathcal{I}_1 \) [44, Section 3]. Without loss of generality, let \( \mathcal{I}_1 := \mathbb{Z}[1, N] \) for ease of notation. We refer to [46] for a discussion on the role of \( \mathcal{I}_0 \) and \( \mathcal{I}_2 \) in terms of closed-loop performance.

\footnote{Random solutions, both \( \phi(k; x, u, \cdot) \) and \( \phi_n(k; x, \cdot) \), exist under the assumption that for all \( x \in \mathbb{R}^n \), the mapping \( \delta \mapsto f(x, \delta) + g(x, \delta) \) is measurable and that \( \kappa \) is measurable, see [45, Section 5.2] and Appendix C for technical details.}
In particular, later on we show that our main results of Section III-A are directly applicable because the sampled nonlinear MPC program $\text{SP}^{\text{MPC}}$ in (29) has non-convex cost, due to the nonlinear dynamics in (26), and convex first-stage constraint. Since the program in (29) is non-convex, and hence the global optimizer is in general not computable exactly, we adopt the following set-based definition of probability of violation.

**Definition 2** (First-stage probability of violation). For given $x \in \mathbb{X}$ and $\mathbb{U}_0 \subseteq \mathbb{U}$, the first-stage probability of violation is given by

$$V^{\text{MPC}}(x, \mathbb{U}_0) := \sup_{u \in \mathbb{U}_0} \mathbb{P}(\{v \in \mathbb{V} \mid f(x, v) + g(x, v)u \notin \mathbb{X}\}).$$

Analogously to Section III-A, see Remark 2, we then consider $M$ directions $c_1, c_2, ..., c_M \in \mathbb{R}^m$, and an arbitrary $\hat{u}_0 \in \mathbb{R}^m$. For instance, but not necessarily, $\hat{u}_0$ may be a known robustly feasible solution. For all $j \in \mathbb{Z}[1, M]$, we define the following SP, where $\bar{v}_0 := (\bar{v}_0^{(1)}, ..., \bar{v}_0^{(N)})$.

$$\text{SP}^j_{1}[\bar{v}_0] : \begin{cases} \min_{\lambda \in \mathbb{R}} & -\lambda \\ \text{sub. to:} & f(x, \bar{v}_0^{(i)}) + g(x, \bar{v}_0^{(i)})(\hat{u}_0 + \lambda c_j) \in \mathbb{X} \quad \forall i \in \mathbb{Z}[1, N] \\ & \hat{u}_0 + \lambda c_j \in \mathbb{U}. \end{cases}$$

Let $\lambda^*_j$ be the optimizer mapping of $\text{SP}^j_{1}$. If $\text{SP}^j_{1}[\bar{v}_0]$ is not feasible, we use the convention that $\lambda^*_j(\bar{x}) := \emptyset$. For all the feasible problems $\text{SP}^j_{1}[\bar{v}_0]$, we define

$$\mathbb{U}_M(\bar{v}_0) := \text{conv}\{\hat{u}_0 + \lambda_1(\bar{v}_0)d_1, \hat{u}_0 + \lambda_2(\bar{v}_0)d_2, ..., \hat{u}_0 + \lambda_M(\bar{v}_0)d_M\}.$$ 

Finally, we solve the following approximation of $\text{SP}^{\text{MPC}}$ in (29).

$$\tilde{\text{SP}}^{\text{MPC}}[\bar{v}^{(1)}, \bar{v}^{(2)}, ...] : \begin{cases} \min_{u \in \mathbb{U}^N} & \sum_{i \in I_0} J(x, u, \bar{v}^{(i)}) \\ \text{sub. to:} & u_0 \in \mathbb{U}_M(\bar{v}_0) \\ & \phi(k; x, u, \bar{v}^{(i)}) \in \mathbb{X} \quad \forall i \in I_2, \forall k \in \mathbb{Z}[2, K] \end{cases}$$

We can now characterize the required sample complexity for the probability of violation to be, with high confidence, below the desired level.

**Theorem 4.** For all $x \in \mathbb{X}$ and $j \in \mathbb{Z}[1, M]$, let $\lambda^*_j$ be the optimizer mapping of $\text{SP}^j_{1}$ in (30), let $\mathbb{U}_M$ be as in (31), and $\epsilon, \beta \in (0, 1)$. Then

$$\mathbb{P}^N(\{v_0 \in \mathbb{V}^N \mid V^{\text{MPC}}(x, \mathbb{U}_M(v_0)) > \epsilon\}) \leq M\Phi\left(\frac{\epsilon}{\min\{m + 1, M\}}, 1, N\right).$$
Consequently, if
\[ N \geq \frac{e}{e-1} \min\{m+1,M\} \ln \left( \frac{M}{\beta} \right), \tag{34} \]
then \( \mathbb{P}^N \left( \{ v_0 \in V_N \mid V_{\text{MPC}}(x, U_M(v_0)) \leq \epsilon \} \right) \geq 1 - \beta, \) i.e., with probability no smaller than \( 1 - \beta, \) any feasible solution of \( \text{SP}_{\text{MPC}}^{(32)} \) satisfies the first state constraint in \( (28). \)

The result of Theorem 4 can be exploited to characterize the expected closed-loop constraint violation as in [44, Theorem 14], under the following assumption [44, Assumption 5].

**Assumption 3 (Recursive feasibility).** \( \text{SP}_{\text{MPC}} \) in (29) admits a feasible solution at every time step almost surely.

**Corollary 4.** Suppose Assumption 3 holds. For all \( x \in X \) and \( v := (v^{(1)}, ..., v^{(N)}) \in V^{KN}, \)
let \( u(x) := (u_0(x), ..., u_{K-1}(x)) \) be any feasible solution of \( \text{SP}_{\text{MPC}}^{(32)} \) in (32), and define \( \kappa(x) := u_0(x). \)
Let \( U_M(k;v) \) be the set \( U_M(v_0) \) in (31) with \( \phi_{\kappa}(k;x,v) \) in place of \( x. \) If \( N \) satisfies
\[ \int_0^1 M\Phi \left( \frac{\nu}{\min\{m+1,M\}}, 1, N \right) d\nu \leq \epsilon, \tag{35} \]
then\(^4\) for all \( k \geq 0 \) it holds that
\[ \mathbb{E} \left[ V_{\text{MPC}}(\phi_{\kappa}(k;x, \cdot), U_M(k; \cdot)) \right] := \int_{V^{(KN+1)_k}} V_{\text{MPC}}\left( \phi_{\kappa}(k;x,v), U_M(k;v) \right) \mathbb{P}^{(KN+1)_k}(dv) \leq \epsilon. \]

The meaning of Corollary 4 is that the expected closed-loop constraint violation, which can be also interpreted as time-average closed-loop constraint violation [46, Section 2.1], is upper bounded by the specified tolerance \( \epsilon \) whenever the sample size \( N \) satisfies (35). A similar result was recently shown in [44, Section 4.2] for uncertain linear systems and hence here extended to the class of uncertain nonlinear control-affine systems in (26).

Numerical simulations of the proposed stochastic nonlinear MPC approach are provided in [46] for a nonholonomic control-affine system, and the benefits with respect to stochastic linear MPC are shown therein.

\(^4\)In [44, Definition 12], a sample size \( N \) is called admissible if it satisfies \( \int_0^1 \Phi(\nu, m, N) d\nu \leq \epsilon, \) which is the counterpart of (35) for random convex programs. For given \( \epsilon \in (0,1), m, M > 0, \) an admissible \( K \) satisfying (35) can be evaluated via a numerical one-dimensional integration.
B. Other non-convex control-design problems

Our scenario approach is suitable for many non-convex control-design problems, such as robust analysis and control synthesis \[26\], \[31\]. In particular, in \[47\] we address control-design via uncertain Bilinear Matrix Inequalities (BMIs) making comparison with the sample complexity based on statistical learning theory, recently derived in \[48\]. Many practical control problems also rely on an uncertain non-convex optimization, for instance reserve scheduling of systems with high wind power generation \[49\], aerospace control \[50\], truss structures \[51\]. Other non-convex control problems that can be addressed via randomization arise in the control of switched systems \[52\], network control \[53\], fault detection and isolation \[54\].

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. For this class of programs, Helly’s dimension can be unbounded. We have derived probabilistic guarantees for all feasible solutions inside a convex set with a-priori chosen complexity, which affects the sample size merely logarithmically.

Our scenario approach also extends to the case with mixed-integer decision variables. We have applied our scenario approach to randomized Model Predictive Control for nonlinear control-affine systems with chance constraints, and outlined many non-convex control-design problems as potential applications.

Finally, we have addressed the measure-theoretic issues regarding the measurability of the optimal value and optimal solutions of random (convex and non-convex) programs, including the well-definiteness of the probability integrals.

ACKNOWLEDGEMENTS

The authors would like to thank Marco Campi, Simone Garatti, Georg Schildbach for fruitful discussions on related topics. Research partially supported by Swiss Nano-Tera.ch under the project HeatReserves.
Fig. 2. The constraints of the problem $\text{SP}_{\text{ex}}[\bar{\omega}]$ with $N = 5$ are represented. The blue surface is the set of points such that $z = -\sqrt{x^2 + y^2}$, while the red hyperplanes are the sets of points such that $z = \cos(\bar{\delta}(i))x + \sin(\bar{\delta}(i))y - 1$, for $i = 1, 2, ..., 5$. The feasible set is the region above the plotted surfaces and the minimization direction is the vertical one, pointing down.

APPENDIX A

COUNTEREXAMPLE WITH UNBOUNDED NUMBER OF SUPPORT CONSTRAINTS

We present an SP, derived from a CCP of the form (2), in which Helly’s dimension [20, Definition 3.1] cannot be bounded. Namely, the number of constraints (“support constraints” [20, Definition 2.1]) needed to characterize the global optimal value equals the number $N$ of samples.

$$\text{SP}_{\text{ex}}[\bar{\omega}] : \begin{cases} 
\min_{(x,y,z) \in \mathbb{R}^3} z \\
\text{sub. to: } z \geq -\sqrt{x^2 + y^2} \\
z \geq \cos(\bar{\delta}(i))x + \sin(\bar{\delta}(i))y - 1 \ \forall i \in \mathbb{Z}[1,N].
\end{cases} \quad (36)$$

The problem can be also written in the form (4), with non-convex cost $J(x, y) := -\sqrt{x^2 + y^2}$ and non-convex constraints $-\sqrt{x^2 + y^2} \geq \cos(\bar{\delta}(i))x + \sin(\bar{\delta}(i))y - 1$. We use the form in (36) to visualize the optimizing direction $-z$, as shown in Figure 2.

Let the drawn samples be $\bar{\delta}(i) = (i - 1) \frac{2\pi}{N}$, for $i = 1, 2, ..., N$. Namely, we divide the $2\pi$-angle into $N$ parts, so that $\bar{\delta}(1) = 0$ and $\bar{\delta}(i+1) = \bar{\delta}(i) + \frac{2\pi}{N}$ for all $i \in \mathbb{Z}[1,N-1]$. We take $N \geq 5$ as $\frac{2\pi}{N} \in (0, \pi/2)$ simplifies the analysis.

We show that all the sampled constraints $z \geq \cos(\bar{\delta}(i))x + \sin(\bar{\delta}(i))y$, for $i = 1, 2, ..., N$, are support constraints, making it impossible to bound Helly’s dimension by some $\zeta < N$. 

January 14, 2014 DRAFT
We first compute the optimal value $J_{ex}^*[\bar{\omega}]$ of $\text{SP}_{ex}[\bar{\omega}]$ in (36). By symmetry and regularity arguments (i.e. continuity of both the objective function and the constraints in the decision variable), an optimizer $(x_N^*, y_N^*, z_N^*)$ can be computed as the intersection of any two adjacent hyperplanes, say $\{(x, y, z) \in \mathbb{R}^3 \mid z = \cos(\tilde{\delta}(i))x + \sin(\tilde{\delta}(i))y - 1\}$ for $i = 1, 2$, and the surface $\{(x, y, z) \in \mathbb{R}^3 \mid z = -\sqrt{x^2 + y^2}\}$. Since $\tilde{\delta}(1) = 0$ and $\tilde{\delta}(2) = \frac{2\pi}{N} =: \theta_N \in (0, \pi/2)$, the optimal value and an optimizer can be computed by solving the system of equations:

$$z = -\sqrt{x^2 + y^2} = x - 1 = \cos(\theta_N)x + \sin(\theta_N)y - 1. \quad (37)$$

From the second and the third equations of (37), we get that $y = \frac{\sin(\theta_N)}{1 + \cos(\theta_N)}x$ and hence from the first equation of (37) we finally get: $\left(\frac{\sin(\theta_N)}{1 + \cos(\theta_N)}\right)^2x^2 + 2x - 1 = 0$. Therefore an optimizer is

$$x_N^* = \sqrt{1 + \left(\frac{\sin(\theta_N)}{1 + \cos(\theta_N)}\right)^2} - 1, \quad y_N^* = \frac{\sin(\theta_N)}{1 + \cos(\theta_N)}x_N^*, \quad z_N^* = x_N^* - 1 \quad (38)$$

and the optimal cost is $J_{ex}^*[\bar{\omega}] = z_N^*$.

We then remove the sample $\tilde{\delta}(2) = \frac{2\pi}{N}$, and hence consider the problem $\text{SP}_{ex}[\bar{\omega} \setminus \tilde{\delta}(2)]$. The optimizer is now unique and lies in the intersection of the hyperplanes $\{(x, y, z) \in \mathbb{R}^3 \mid z = \cos(\tilde{\delta}(i))x + \sin(\tilde{\delta}(i))y - 1\}$, for $i = 1, 3$, and the surface $\{(x, y, z) \in \mathbb{R}^3 \mid z = -\sqrt{x^2 + y^2}\}$. We just need to solve the system of equations (37), but with $\tilde{\delta}(3) := 2\theta_N = \frac{4\pi}{N}$ in place of $\theta_N$ in the third equation. Therefore we obtain almost the same solution in (37), but with $2\theta_N$ in place of $\theta_N$. Since the optimal cost

$$J_{ex}^*[\bar{\omega} \setminus \tilde{\delta}(2)] = \sqrt{1 + \left(\frac{\sin(2\theta_N)}{1 + \cos(2\theta_N)}\right)^2} - 1$$

is strictly smaller than $J_{ex}^*[\bar{\omega}]$ (as $x_N^* < x_{N+1}^*$ for all $N \geq 5$), it follows that the constraint associated with $\tilde{\delta}(2)$ is a support constraint. Figure 3 shows the optimizer of the problem $\text{SP}_{ex}[\bar{\omega} \setminus \{\tilde{\delta}(2)\}]$. Because of the symmetry of the problem with respect to rotations around the $z$-axis, we conclude that all the $N$ affine constraints $z \geq \cos(\tilde{\delta}(i))x + \sin(\tilde{\delta}(i))y - 1$, for $i = 1, 2, \ldots, N$, are support constraints as well, i.e. $J_{ex}^*[\bar{\omega} \setminus \tilde{\delta}(i)] < J_{ex}^*[\bar{\omega}]$ for all $i \in \mathbb{Z}[1, N]$. This proves that Helly’s dimension cannot be upper bounded by some a-priori fixed $\zeta < N$.

Finally, in view of [19 Theorem 1], it suffices to find at least one probability measure so that the extraction of the above samples $\tilde{\delta}(1), \tilde{\delta}(2), \ldots, \tilde{\delta}(N)$ happens with non-zero probability. For
The constraints of the problem $SP_{\alpha}[\bar{\omega} \setminus \bar{\delta}^{(2)}]$ with $N = 5$ are represented. The feasible set is the region above the blue surface, which is the set of points such that $z = \max_{i \in \{1, 3, 4, 5\}} \{-\sqrt{x^2 + y^2}, \cos(\bar{\delta}(i))x + \sin(\bar{\delta}(i))y - 1\}$. The red dot represents the optimizer $(x^*, y^*, z^*)$ which has a cost $J(x^*, y^*, z^*) = J^{*}[\bar{\omega} \setminus \bar{\delta}^{(2)}] < J^{*}[\bar{\omega}]$.

instance, this holds true if $P$ is such that $P(\{\bar{\delta}(i)\}) = 1/N$ for all $i \in \mathbb{Z}[1, N]$. Moreover, it is also possible to have a distribution about the above points $\bar{\delta}(1), \bar{\delta}(2), \ldots, \bar{\delta}(N)$ that has a density, but is narrow enough to preserve the property that $J^{*}[\bar{\omega} \setminus \bar{\delta}^{(2)}] < J^{*}[\bar{\omega}]$.

APPENDIX B
PROOFS

Proof of Theorem 1

Let $\mathcal{X}_\epsilon := \{x \in \mathcal{X} \mid P(\{\delta \in \Delta \mid g(x, \delta) \leq 0\}) \geq 1 - \epsilon\}$ be the feasibility set of CCP($\epsilon$) in (1). Take any arbitrary $y \in \text{conv}(\mathcal{X}_\epsilon)$. It follows from Caratheodory’s Theorem [55, Theorem 17.1] that there exist $x_1, x_2, ..., x_{n+1} \in \mathcal{X}_\epsilon$ such that $y \in \text{conv}(\{x_1, x_2, ..., x_{n+1}\})$, i.e. $y = \sum_{i=1}^{n+1} \alpha_i x_i$ 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.

$$
P(\{\delta \in \Delta \mid g(y, \delta) > 0\}) = P(\{\delta \in \Delta \mid g(\sum_{i=1}^{n+1} \alpha_i x_i, \delta) > 0\}) 
\leq P(\{\delta \in \Delta \mid \sum_{i=1}^{n+1} \alpha_i g(x_i, \delta) > 0\}) 
\leq P(\{\delta \in \Delta \mid \max_{i \in [1, n+1]} \alpha_i g(x_i, \delta) > 0\}) 
= P(\bigcup_{i=1}^{n+1} \{\delta \in \Delta \mid g(x_i, \delta) > 0\}) 
\leq \sum_{i=1}^{n+1} P(\{\delta \in \Delta \mid g(x_i, \delta) > 0\}) 
\leq (n + 1)\epsilon.
$$

(39)

The last inequality follows from the fact that $x_1, x_2, ..., x_{n+1} \in \mathcal{X}_\epsilon$. 

January 14, 2014
Since $y \in \text{conv}(\mathcal{X}_c)$ has been chosen arbitrarily, it follows that $V(\text{conv}(\mathcal{X}_c)) \leq (n + 1)\epsilon$. ■

Proof of Lemma 1

\[ P_N \left( \left\{ \omega \in \Delta^N \mid V(\{x^*_1(\omega), \ldots, x^*_M(\omega)\}) > \epsilon \right\} \right) = P_N \left( \bigcup_{j=1}^M \left\{ \omega \in \Delta^N \mid V(\{x^*_j(\omega)\}) > \epsilon \right\} \right) \leq \sum_{k=1}^M P_N \left( \left\{ \omega \in \Delta^N \mid V(\{x^*_k(\omega)\}) > \epsilon \right\} \right) \leq \sum_{k=1}^M \beta_k, \] where the last inequality follows from Assumption 1. ■

Proof of Theorem 2

For all $\omega \in \Delta^N$, from the definition of the supremum $V(\mathcal{X}) = \sup_{x \in \mathcal{X}_M(\omega)} V(\{x\})$ it holds that for all $\epsilon' > 0$ there exists $\xi^*_M(\omega) \in \mathcal{X}_M(\omega) = \text{conv}(\{x^*_1(\omega), x^*_2(\omega), \ldots, x^*_M(\omega)\})$ such that

\[ V(\mathcal{X}_M(\omega)) = \sup_{x \in \mathcal{X}_M(\omega)} V(\{x\}) < V(\xi^*_M(\omega)) + \epsilon'. \] (40)

Now, for all $\omega \in \Delta^N$, we denote by $\mathcal{I}(\omega) \subset \mathbb{Z}[1, M]$ the set of indices of cardinality $|\mathcal{I}(\omega)| = \min\{n + 1, M\}$, with “minimum lexicographic order” such that we have the inclusion $\xi^*_M(\omega) \in \text{conv}(\{x^*_j(\omega) \mid j \in \mathcal{I}(\omega)\})$. Since $\mathcal{X}_M(\omega)$ is convex and compact, it follows from Caratheodory’s Theorem [55, Theorem 17.1] that such a set of indices $\mathcal{I}(\omega)$ always exists. It also follows that there exists a unique set of coefficients $\alpha_1(\omega), \alpha_2(\omega), \ldots, \alpha_{n+1}(\omega) \in [0, 1]$ such that

\[ \sum_{j \in \mathcal{I}(\omega)} \alpha_j(\omega) = 1 \quad \text{and} \quad \xi^*_M(\omega) = \sum_{j \in \mathcal{I}(\omega)} \alpha_j(\omega)x^*_j(\omega). \] (41)

In the following inequalities, we exploit (40), (41) and the convexity of the mapping $x \mapsto g(x, \delta)$ for each fixed $\delta \in \Delta$ from Standing Assumption 1, and we can take $\epsilon' \in (0, \epsilon)$ without

\[^5\text{With “minimum lexicographic order” we mean the following ordering: } \{i_1, i_2, \ldots, i_n\} < \{j_1, j_2, \ldots, j_n\} \text{ if there exists } k \in \mathbb{Z}[1, n] \text{ such that } i_1 = j_1, \ldots, i_{k-1} = j_{k-1}, \text{ and } i_k < j_k. \]
loss of generality.
\[
\mathbb{P}^n \left( \{ \omega \in \Delta^N \mid V(X_M(\omega)) > \epsilon \} \right) = \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid \sup_{x \in X_M(\omega)} V(\{x\}) > \epsilon \} \right)
\]
\[
\leq \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid V(\{\xi^*_\omega\}) > \epsilon - \epsilon' \} \right)
\]
\[
= \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid \mathbb{P}\left( \left\{ \delta \in \Delta \mid \sum_{j \in I(\omega)} \alpha_j(\omega) x_j^*(\omega, \delta) > 0 \right\} \right) > \epsilon - \epsilon' \} \right)
\]
\[
\leq \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid \mathbb{P}\left( \left\{ \delta \in \Delta \mid \sum_{j \in I(\omega)} \alpha_j(\omega) g(x_j^*(\omega, \delta)) > 0 \right\} \right) > \epsilon - \epsilon' \} \right)
\]
\[
\leq \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid \max_{j \in I(\omega)} g(x_j^*(\omega, \delta)) > \epsilon - \epsilon' \} \right)
\]
\[
= \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid \mathbb{P}\left( \left\{ \delta \in \Delta \mid g(x_j^*(\omega, \delta)) > 0 \right\} \right) > \epsilon - \epsilon' \} \right)
\]
\[
\leq \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid V(\{x_j^*(\omega) \mid j \in I(\omega)\}) > \epsilon - \epsilon' \} \right)
\]
\[
\leq \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid V(\{x_1^*(\omega), x_2^*(\omega), ..., x_M^*(\omega)\}) > \epsilon - \epsilon' \} \right)
\]

(42)

Since for all \( k \in \mathbb{Z}[1, M] \), \( x_k^*(\cdot) \) is the optimizer mapping of \( SP_1[\cdot] \) in (8), from Theorem 1], \( \beta_1 = \Phi(\frac{-\epsilon'}{n+1}, n, N) \) and so that, for all \( \epsilon' > 0 \), we get
\[
\mathbb{P}^n \left( \{ \omega \in \Delta^N \mid V(X_M(\omega)) > \epsilon \} \right) \leq \mathbb{P}^n \left( \{ \omega \in \Delta^N \mid V(\{x_1^*(\omega), x_2^*(\omega), ..., x_M^*(\omega)\}) > \epsilon - \epsilon' \} \right) \leq M\Phi(\frac{-\epsilon'}{n+1}, n, N)
\]

Then, since for all \( n, N \geq 1 \) the mapping \( \epsilon \mapsto \Phi(\epsilon, n, N) \) is continuous, we have that
\[
\lim_{\epsilon' \to 0} M\Phi(\frac{-\epsilon'}{n+1}, n, N) = \lim_{\epsilon' \to 0} M\Phi(\frac{-\epsilon'}{n+1}, n, N) = M\Phi(\frac{-\epsilon}{n+1}, n, N), \text{ which proves (11)}.
\]

**Proof of Corollary 1**

It follows from Carathéodory’s Theorem [55, Theorem 17.1] that, for each \( \omega \in \Delta^N \), there exist the sets \( X_M^i(\omega) := \text{conv}(\{x_k^*(\omega) \mid k \in I_i\}) \), for \( i = 1, 2, ..., (\frac{M}{n+1}) \), where each \( I_i \) is a set of indices of cardinality \( n + 1 \), such that \( X_M(\omega) = \bigcup_{i=1}^{\frac{M}{n+1}} X_M^i(\omega) \). Therefore we can write
\[
\mathbb{P}^n \left( \{ \omega \in \Delta^N \mid \sup_{x \in X_M(\omega)} V(\{x\}) > \epsilon \} \right)
\]
\[
= \mathbb{P}^n \left( \left\{ \omega \in \Delta^N \mid \max_{i \in [1, (\frac{M}{n+1})]} \sup_{x \in X_M^i(\omega)} V(\{x\}) > \epsilon \right\} \right)
\]
\[
= \mathbb{P}^n \left( \bigcup_{i=1}^{\frac{M}{n+1}} \left\{ \omega \in \Delta^N \mid \sup_{x \in X_M^i(\omega)} V(\{x\}) > \epsilon \right\} \right)
\]
\[
\leq \sum_{i=1}^{\frac{M}{n+1}} \mathbb{P}^n \left( \left\{ \omega \in \Delta^N \mid \sup_{x \in X_M^i(\omega)} V(\{x\}) > \epsilon \right\} \right)
\]
\[
\leq \left( \frac{M}{n+1} \right) \mathbb{P}^n \left( \left\{ \omega \in \Delta^N \mid \sup_{x \in X_M^i(\omega)} V(\{x\}) > \epsilon \right\} \right),
\]
where in the last inequality we consider the first set of indices without loss of generality, similarly to [20] Proof of Theorem 3.3, pag. 3435. It follows from Theorem 2 and [35] Equation (24) that for all $\epsilon' > 0$ there exists $\xi^*(\omega) \in \mathbb{X}^{(1)}_M(\omega) = \text{conv} \left( \left\{ x^*_1(\omega), x^*_2(\omega), \ldots, x^*_{n+1}(\omega) \right\} \right)$ such that

$$
\mathbb{P}^N \left( \left\{ \omega \in \Delta^N \mid V(\mathbb{X}_M(\omega)) > \epsilon \right\} \right) \leq \binom{M}{n+1} \mathbb{P}^N \left( \left\{ \omega \in \Delta^N \mid V(\{\xi^*(\omega)\}) > \epsilon - \epsilon' \right\} \right) \leq \left( \frac{M}{n+1} \right) \Phi (\epsilon - \epsilon', \zeta(n+1), N)
$$

and hence, after taking the $\lim \sup \epsilon' \to 0$ on both sides of the inequality, we finally get the inequality

$$
\mathbb{P}^N \left( \left\{ \omega \in \Delta^N \mid V(\mathbb{X}_M(\omega)) > \epsilon \right\} \right) \leq \left( \frac{M}{n+1} \right) \Phi (\epsilon, \zeta(n+1), N).
$$

\[\square\]

**Proof of Corollary 2**

If follows from (11) that we need to find $N$ such that $\Phi \left( \frac{1}{n+1}, n, N \right) < \beta / M$. The proof follows similarly to [31] Proof of Theorem 3].

\[\square\]

**Proof of Theorem 3**

The proof is similar to the proof of Theorem 2

\[\square\]

**Proof of Corollary 3**

The proof is similar to the proof of Corollary 2

\[\square\]

**Proof of Theorem 4**

For each $j \in \mathbb{Z}[1, M]$, we consider the random convex problem $\text{SP}^1_j[\cdot]$ in (30), with unique optimizer mapping $\lambda^*_j(\cdot)$. Since the dimension of the decision variable is 1, i.e. $u^*_j(\cdot) := \hat{u}_0 + \lambda^*_j(\cdot)c_j$, it follows from [19] Theorem 1, [20] Theorem 3.3 that, for all $j \in \mathbb{Z}[1, M]$, we have

$$
\mathbb{P}^N \left( \left\{ \omega \in \mathcal{V}^N \mid V^{\text{MPC}} \left( \{u^*_j(\omega)\} \right) > \epsilon \right\} \right) \leq \Phi(\epsilon, 1, N).
$$

Then, from Lemma 1 we have that:

$$
\mathbb{P}^N \left( \left\{ \omega \in \mathcal{V}^N \mid V^{\text{MPC}} \left( \{u^*_1(\omega), u^*_2(\omega), \ldots, u^*_M(\omega)\} \right) > \epsilon \right\} \right) \leq M \Phi(\epsilon, 1, N).
$$
We now notice that the CCP in (28) is of the same form of (2), with the constraints
\[ P^k \left( \{ v \in V^k \mid \phi(k; x, \cdot, v) \notin X \} \right) \leq \epsilon, \text{ for } k \geq 2, \text{ in place of } h(\cdot) \leq 0. \]Therefore to conclude the proof we just have to follow the steps of Remark 1 and the proof of Theorem 2 with \( \{ u^*_1(\omega), \ldots u^*_M(\omega) \} \) in place of \( \{ x^*_1(\omega), \ldots x^*_M(\omega) \} \), and finally derive the sample size \( N \) according to (35).

\[ \blacksquare \]

Proof of Corollary 4

Since the sample size \( N \) satisfies (35), the proof follows from [44, Section 4.2].

\[ \blacksquare \]

APPENDIX C

MEASURABILITY OF OPTIMAL VALUE AND OF OPTIMAL SOLUTIONS

In this section, we adopt the following notion of measurability from [45, Section 2]. Let \( B(\mathbb{R}^n) \) denote the Borel field, the subsets of \( \mathbb{R}^n \) generated from open subsets of \( \mathbb{R}^n \) through complements and finite countable unions. A set \( F \subset \mathbb{R}^n \) is measurable if \( F \in B(\mathbb{R}^n) \). A set-valued mapping \( M : \mathbb{R}^n \Rightarrow \mathbb{R}^m \) is measurable [30, Definition 14.1] if for each open set \( O \subset \mathbb{R}^m \) the set \( M^{-1}(O) := \{ v \in \mathbb{R}^n \mid M(v) \cap O \neq \emptyset \} \) is measurable. When the values of \( M \) are closed, measurability is equivalent to \( M^{-1}(C) \) being measurable for each closed set \( C \in \mathbb{R}^m \) [30, Theorem 14.3]. Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space, where \( \mathbb{P} \) is a probability measure on \( \mathbb{R}^n \). A set \( F \subset \mathbb{R}^n \) is universally measurable if it belongs to the Lebesgue completion of \( B(\mathbb{R}^n) \). A set-valued mapping \( M : \mathbb{R}^n \Rightarrow \mathbb{R}^m \) is universally measurable if the set \( M^{-1}(S) \) is universally measurable for all \( S \in B(\mathbb{R}^m) \) [56, Section 7.1, pag. 68]. If \( \varphi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{ \pm \infty \} \) is a (universally) measurable function, then the integral \( I[\varphi] := \int_{\mathbb{R}^m} \varphi(\omega) \mathbb{P}(d\omega) \) is (nearly) well defined [30, Chapter 14, pag. 643].

The following result shows (near) well definiteness of the stated probability integrals.

**Theorem 5.** For all \( x \in \mathcal{X} \), the probability integral \( \mathbb{P} \left( \{ \delta \in \Delta \mid g(x, \delta) \leq 0 \} \right) \) is well defined. For any measurable set-valued mapping \( X : \Delta^N \Rightarrow \mathbb{R}^n \) and \( \epsilon \in (0, 1) \), the probability integral \( \mathbb{P}^N \left( \{ \omega \in \Delta^N \mid V(X(\omega)) > \epsilon \} \right) \) is nearly well defined. \( \square \)

**Proof:** From Standing Assumption 1 we have that \( g \) is a lower semicontinuous convex integrand, and hence a normal integrand [30, Proposition 14.39]. Therefore, for all \( x \in \mathcal{X} \), the
set \( \{ \delta \in \Delta \mid g(x, \delta) \leq 0 \} \) is measurable \cite{30} Proposition 14.33] and in turn the probabilistic measure \( \mathbb{P}(\{ \delta \in \Delta \mid g(x, \delta) \leq 0 \}) \) is well defined.

For the second statement, we show that, for all set-valued measurable mappings \( \mathbb{X} \), the mapping \( \omega \mapsto \sup_{x \in \mathbb{X}(\omega)} V(\{ x \}) \) is universally measurable. Since \( g \) is a normal integrand, for any finite non-negative measure \( \mu \) on \( \mathcal{X} \subseteq \mathbb{R}^n \), we have that \( g \) is jointly \((\mathbb{P} \otimes \mu)\)-measurable \cite{30} Corollary 14.34]. Indeed, the set \( \mathcal{A} := \{(x, \delta) \in \mathcal{X} \times \Delta \mid g(x, \delta) \leq 0 \} \) is \((\mathbb{P} \otimes \mu)\)-measurable \cite{30} Proof of Corollary 14.34], and in turn the mapping \( (x, \delta) \mapsto \mathbb{1}_{\mathcal{A}}(x, \delta) \) is measurable. It then follows from Fubini’s Theorem \cite{57} Theorem 8.8 (a)] that the mapping \( x \mapsto \int_{\Delta} \mathbb{1}_{\mathcal{A}}(x, \delta) \mathbb{P}(d\delta) = \mathbb{P}(\{ \delta \in \Delta \mid g(x, \delta) \leq 0 \}) \) is measurable, and in turn \( x \mapsto V(\{ x \}) = 1 - \mathbb{P}(\{ \delta \in \Delta \mid g(x, \delta) \leq 0 \}) \) is measurable as well \cite{30} Proposition 14.11 (c)]. Since \( V \) is measurable, it follows from \cite{58} Theorem 2.17 (a)] that \( \omega \mapsto \sup_{x \in \mathbb{X}(\omega)} V(\{ x \}) \) is analytic and hence universally measurable \cite{58} Fact 2.9].}

\textbf{Remark 3.} According to the proof of Theorem \cite{5} the mapping \( \omega \mapsto \bar{V}(\mathbb{X}(\omega)) \) is not measurable, but only nearly measurable. However, near measurability is sufficient for the purposes of most applications, for instance in game-theory and econometrics, see \cite{58} and the references therein.

Notice however that the upper closure of \( V \), i.e. \( \bar{V}(\{ x \}) := \limsup_{y \to x} V(\{ y \}) \), is such that the integral \( \mathbb{P}^N(\{ \omega \in \Delta^N \mid \bar{V}(\mathbb{X}(\omega)) > \epsilon \}) \) is well defined. In fact, since \( \bar{V} \) is upper semicontinuous by construction, \( -\bar{V} \) is an autonomous, lower semicontinuous, normal integrand \cite{30} Example 14.30]. Then it follows from \cite{30} Example 14.32, Theorem 14.37] that the mapping \( \omega \mapsto \bar{V}(\mathbb{X}(\omega)) := \sup_{x \in \mathbb{X}(\omega)} \bar{V}(\{ x \}) \) is measurable.

We can now show the following result on the measurability of optimal value and of optimal solutions of \( \text{SP}[\omega] \) in \cite{4}, which means that they actually are random variables.

\textbf{Theorem 6.} Let \( J^* : \Delta^N \to \mathbb{R} \) and \( \mathcal{X}^* : \Delta^N \Rightarrow \mathcal{X} \) be the mappings such that, for all \( \omega \in \Delta^N \), \( J^*(\omega) \) and \( \mathcal{X}^*(\omega) \) are, respectively, the optimal value and the set of optimizers of \( \text{SP}[\omega] \) in \cite{4}. Then \( J^* \) is measurable, and \( \mathcal{X}^* \) is closed-valued and measurable. Moreover, \( \mathcal{X}^* \) admits a measurable selection, i.e., there exists a measurable mapping \( x^* : \Delta^N \to \mathcal{X} \) such that \( x^*(\omega) \in \mathcal{X}^*(\omega) \) for all \( \omega \in \Delta^N \). \hfill \Box

\textbf{Proof:} Since the mapping \( x \mapsto g(x, \delta) \) is convex and lower semicontinuous for each \( \delta \), and the mapping \( \delta \mapsto g(x, \delta) \) is measurable for each \( x \), we have that \( g \) is a lower semicontinuous
integrand and hence a normal integrand [30, Definition 14.27, Proposition 14.39]. For all \( i \in \mathbb{Z}[1, N] \), we consider the lower semicontinuous convex, and hence normal [30, Proposition 14.39], integrand \( g_i : \mathcal{X} \times \Delta^N \to \mathbb{R} \) defined as \( g_i(x, \omega) = g_i(x, (\delta_1, \delta_2, ..., \delta_N)) := g(x, \delta_i) \). Then we consider the mapping \( \bar{g} : \mathcal{X} \times \Delta^N \to \mathbb{R} \) defined as \( \bar{g}(x, \omega) := \max_{i \in \mathbb{Z}[1, N]} g_i(x, \omega) \), which is a normal integrand because the pointwise maximum of the normal integrands \( g_1, g_2, ..., g_N \) [30, Proposition 14.39 (a)]. We now consider the set-valued mapping \( C : \Delta^n \to \mathcal{X} \) defined as \( C(\omega) := \{ x \in \mathcal{X} \mid \bar{g}(x, \omega) \leq 0 \} \). Since \( \bar{g} \) is a normal integrand, it follows from [30, Proposition 14.33] that the level-set mapping \( C \) is closed-valued and measurable. Thus, we can define the indicator integrand \( \mathbb{1}_C : \mathcal{X} \times \Delta^N \to \{ 0, \infty \} \) as \( \mathbb{1}_C(x, \omega) = \mathbb{1}_{C(\omega)}(x) := \{ 0 \text{ if } x \in C(\omega), \infty \text{ otherwise} \} \). Since \( C \) is closed-valued and measurable, the mapping \( \mathbb{1}_C \) is a normal integrand [30, Example 14.32]. Now, the problem \( \text{SP}[\omega] \) in (4) can be written as \( \min_{x \in \mathcal{X}} c^\top x \) sub. to \( x \in C(\omega) \), which is equivalent [30, Section 1.A] to \( \min_{x \in \mathbb{R}^n} J(x) + \mathbb{1}_C(x, \omega) \). We notice that the mapping \( (x, \omega) \mapsto \varphi(x, \omega) := J(x) + \mathbb{1}_C(x, \omega) \) is a normal integrand as \( J \) is lower semicontinuous [30, Example 14.30, Example 14.32, Proposition 14.44 (c)]. It finally follows from [30, Theorem 14.37] that the optimal value mapping \( \omega \mapsto J^*(\omega) := \inf_{x \in \mathbb{R}^n} \varphi(x, \omega) \) is measurable; also, the set-valued mapping \( \omega \mapsto \mathcal{X}^*(\omega) := \arg \min_{x \in \mathbb{R}^n} \varphi(x, \omega) \) is closed-valued and measurable. Moreover, the set \( \{ \omega \in \Delta^N \mid \mathcal{X}^*(\omega) \neq \emptyset \} \) is measurable, and it is possible for each \( \omega \in \Delta^N \) to select a minimizing point \( x^*(\omega) \) in such a manner that the mapping \( \omega \mapsto x^*(\omega) \) is measurable [30, Corollary 14.6, Theorem 14.37].

In the following result, we show that if the set of optimizers \( \mathcal{X}^* \) of \( \text{SP} \) in (4) is not a singleton, convex and lower semicontinuous tie-break rules \( \varphi \) are sufficient to guarantee measurability of the optimizer \( x^* \) (whenever it is unique). Applying a tie-break rule \( \varphi \) basically means to solve the following program, where \( J^*(\omega) \) is the optimal value of \( \text{SP}[\omega] \) in (4).

\[
\text{SP}_{t-b}[\omega] : \begin{cases} 
\min_{x \in \mathcal{X}} \varphi(x) \\
\text{sub. to: } \quad g(x, \delta^i) \leq 0 \quad \forall i \in \mathbb{Z}[1, N] \\
\quad J(x) \leq J^*(\omega)
\end{cases}
\] (43)

**Corollary 5.** Let \( \varphi : \mathbb{R}^n \to \mathbb{R} \) be a convex and lower semicontinuous function. Let \( J^* : \Delta^N \to \mathbb{R} \) and \( x^*_{t-b} : \Delta^N \to \mathcal{X} \) be such that, for all \( \omega \in \Delta^N \), \( J^*(\omega) \) and \( x^*_{t-b}(\omega) \) are, respectively, the optimal value of \( \text{SP}[\omega] \) in (4) and the unique optimal solution of \( \text{SP}_{t-b}[\omega] \) in (43). Then \( x^*_{t-b} \) is measurable. \( \square \)
Proof: We first define the normal integrand
\[ \bar{g}(x, \omega) = \bar{g}(x, (\delta_1, \delta_2, ..., \delta_N)) := \max_{i \in \mathbb{Z}[1,N]} g_i(x, \omega) = \max_{i \in \mathbb{Z}[1,N]} g(x, \delta_i), \]
as in the proof of Theorem 6. Since \( J \) is lower semicontinuous, it is an autonomous integrand and hence a normal integrand [30, Example 14.30]; moreover, since \( J^* \) is measurable from Theorem 6, it is a (Carathéodory) normal integrand as well [30, Example 14.29]. Therefore also the mapping \((x, \omega) \mapsto J(x) - J^*(\omega)\) is a normal integrand as well [30 Proposition 14.44 (c)], and in turn, the mapping \( \bar{g}(x, \omega) := \max\{\bar{g}(x, \omega), J(x) - J^*(\omega)\} \) is a normal integrand as well. Then, we can just follow the proof of Theorem 6 with \( \bar{g} \) in place of \( \bar{g} \).

Remark 4. In (43), if \( J \) is convex and \( \varphi \) is strictly convex then an optimal solution \( x_{t,b}^*(\omega) \) of \( \text{SP}_{t,b}[\omega] \) is the unique optimal solution. \( \square \)

We finally mention that the convex hull of measurable singleton mappings is measurable as well, so that \( P_N \left\{ \omega \in \Delta^N \mid V(X_M(\omega)) > \epsilon \right\} \) is well defined from Theorem 5.

Corollary 6. The set-valued mapping \( X_M \) in (9) is measurable. \( \square \)

Proof: According to Theorem 6 and Remark 4, the unique optimal solutions \( x_1^*, x_2^*, ..., x_M^* \), respectively of \( \text{SP}_1, \text{SP}_2, ..., \text{SP}_M \), are measurable mappings. Then the proof directly follows as \( X_M \) is the convex-hull set-valued mapping of a countable union of measurable mappings [30 Proposition 114.11 (b), Example 14.12 (a)]. \( \blacksquare \)

References


