2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

Tight generalization guarantees for the sampling and discarding approach to scenario optimization

Licio Romao, Kostas Margellos and Antonis Papachristodoulou

***Abstract*— We consider the scenario approach theory to deal with convex optimization programs affected by uncertainty, which is in turn represented by means of scenarios. An approach to deal with such programs while trading feasibility to performance is known as sampling and discarding in the scenario approach literature. Existing bounds on the probability of constraint satisfaction for such programs are not tight. In this paper we use learning theoretic concepts based on the notion of compression to show that for a particular class of convex scenario programs, namely, the so called fully-supported ones, and under a particular scenario discarding scheme, a tight bound can be obtained. We illustrate our developments by means of an example that admits an analytic solution.**

I. INTRODUCTION

The scenario approach theory [1]–[5] is a successful randomization technique to solve convex uncertain optimization problems. It is based on representing uncertainty by means of scenarios, accompanying the optimal solution of the resulting scenario program with certificates on the probability of constraint violation. The scenario approach theory was introduced in [6] and has undergone several developments, including tightness of the bound on the probability of constraint violation [2], the sampling and discarding [7] or constraint removal scheme [8] to trade feasibility to performance, extensions to non-convex programs [9], [10] and to min-max [11] problems, and, more recently, *a posteriori* assessing the scenario solution [4], [10], [12], [13], providing game theoretic extensions [14], [15] and an extension applied to the coverage theory for the least square estimate [16].

This paper is related to the *a priori* results of the scenario approach theory that link the confidence, probability of constraint violation, and the number of samples through a simple combinatorial formula that holds for all probability distributions [2], [7], [8] and all convex programs. In the standard formulation of the scenario approach theory one constraint is enforced for each scenario, often leading to conservative values for the optimal objective value, especially when a large amount of scenarios is used. To mitigate this fact, [7], [8] have developed a strategy that allows the decision maker to discard some of the original scenarios. This strategy was introduced in [7] and was termed 'sampling and discarding approach', as well as in [8] where it was termed 'scenario approach with constraint removal'. Other approaches within the scenario theory that can be used to trade probability of constraint violation to performance are presented in [11], [13], where [11] characterizes the joint probability of the risks associated with the empirical costs of a min-max sample-based solution and [13] employs relaxation of the constraints and an *a posteriori* result of the scenario approach theory to assert with high confidence upper and lower bounds on the probability of constraint violation.

The main motivation for the results in this paper is the fact that, unlike the tight bound of [2] that holds with equality for the class of fully-supported convex optimization problems, the bound in [7] is not tight. In fact, it is shown in [7] that there exists an instance of a particular combinatorial constraint removal scheme that yields a solution that possesses a tighter bound on the probability of constraint violation; however, the presented argument is not constructive.

In this paper we revisit the sampling and discarding framework and prove two main results for the class of fully-supported problems: (1) We show that a tighter bound on the probability of constraint violation can be obtained by considering a removal scheme that consists of a cascade of scenario programs, where at each stage the support set, a concept at the core of the scenario approach theory that will be formalized in the sequel, associated to the optimal solution is removed. To establish this tighter bound we analyze the constraint violation properties of the resulting solution within a probably approximately correct (PAC) [17] learning framework based on the notion of compression [18], [19]; (2) Similarly to [2], we show that the proposed bound on the probability of constraint violation is tight by characterizing a class of optimization problems that achieves the proposed bound with equality. It appears that the proposed bound is more general, and albeit not tight, is valid for any convex scenario program under a mild non-degeneracy assumption. This extension can be found in [20].

Section II provides some background concepts from the scenario approach and learning theoretic literature. Section III introduces the proposed discarding scheme and states our main result. Section IV shows that the proposed result is tight for a certain class of programs, and illustrates this by means of an example that admits an analytic solution. Finally, Section V concludes the paper and provides some directions for future work. All omitted proofs can be found in [20].

II. BACKGROUND KNOWLEDGE

A. The scenario theory

Let $(\Delta, \mathcal{F}, \mathbb{P})$ be a probability measure space, where Δ is the domain in which random variables are defined, $\mathcal{F}$ is a σ-algebra of subsets of Δ, and $\mathbb{P} : \Delta \to [0, 1]$ is a probability measure on Δ. We assume that the probability $\mathbb{P}$ is time-invariant and unknown, and that m independent samples from $\mathbb{P}$ are available. We denote these samples as $S =$

L. Romao is supported by the Coordination for the Improvement of Higher Education Personnel (CAPES) - Brazil. The work of K. Margellos and A. Papachristodoulou has been supported by EPSRC UK under grants EP/P03277X/1 and EP/M002454/1, respectively.

The authors are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK. {licio.romao, kostas.margellos, antonis}@eng.ox.ac.uk.

$(\delta_1, \ldots, \delta_m)$, and refer to each element $\delta_i \in S$ as a scenario. Note that S can be seen as an element in the product space $\Delta^m = \prod_{i=1}^m \Delta$ and is distributed according to $\mathbb{P}^m$, due to the fact that δ_i's, $i = 1, \ldots, m$, are assumed to be independent and identically distributed (i.i.d.). In this context, we can associate a useful probability space, namely, $(\Delta^m, \otimes\mathcal{F}, \mathbb{P}^m)$, where $\otimes\mathcal{F}$ is the smallest σ-algebra containing $\prod_{i=1}^m \mathcal{F}$.

Given $m \in \mathbb{N}$, let $S = (\delta_1, \ldots, \delta_m)$ be a collection of m i.i.d. samples as explained in the previous paragraph, and consider the optimization problem

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & g(x, \delta) \le 0, \text{ for all } \delta \in S, \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^d$ is the optimization variable, $c \in \mathbb{R}^d$ is a given vector representing the objective function, $g : \mathbb{R}^d \times \Delta \to \mathbb{R}$ is convex in x for all $\delta \in S$, and $X \subset \mathbb{R}^d$ is a closed convex set. Note that the objective function in (1) can be considered linear without loss of generality, as otherwise one may use an epigraphic formulation to recast any convex problem into the form of (1).

Assumption *1:* For each $m \in \mathbb{N}$, the solution of the optimization problem (1) exists and is unique, and its feasibility region has non-empty interior.

The uniqueness requirement is a mild assumption in most applications and can be ensured by a tie-break rule procedure that selects one solution among all the elements in the optimal set of (1). Similarly, existence can be ensured by standard techniques in variational methods, for instance, the optimal set of (1) is non-empty if we assume that there exists a $\gamma \in \mathbb{R}$ such that the set $X \cap \{x \in \mathbb{R} : c^\top x \le \gamma\}$ is compact or if X is compact.

We refer to problem (1) as a scenario program, as one constraint is enforced for each scenario in S. Let $x^\star(S)$ be the optimal solution of (1), where the argument aims to emphasize the dependence on the original samples given by S. The authors in [2] proved the following result.

Theorem *1 (Theorem 1, [2]):* Consider Assumption 1, and fix $\epsilon \in (0, 1)$. Let $m \in \mathbb{N}$ be given and denote by $x^\star(S)$ the optimal solution of (1). Then we have that

$$\mathbb{P}^m \left\{ (\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta : g(x^\star(S), \delta) > 0\} > \epsilon \right\} \le \sum_{i=0}^{d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{2}$$

The left-hand side of Theorem 1 is composed of two nested probabilities: the outer one represents the confidence with which the bound is valid, and the inner one stands for the risk incurred by the optimal solution of problem (1).

Definition *1 (Support scenario and support set):* An element $\delta \in S = (\delta_1, \ldots, \delta_m)$ is called a support scenario if its removal changes the optimal solution $x^\star(S)$ of problem (1). The support set is the collection of all support scenarios and is denoted by $\mathrm{supp}(x^\star(S))$.

Definition *2 (Fully supported problems):* A scenario program is called fully supported if, with $\mathbb{P}^m$ probability one, the cardinality of the support set is equal to d.

An important feature of the bound in (2) is the fact that it holds for all convex optimization problems and all distributions $\mathbb{P}$. In fact, [2] showed that the class of fully-supported optimization problems achieves the bound of Theorem 1. In this sense, the bound in (2) is said to be tight.

B. The sampling and discarding approach

Let $m \in \mathbb{N}$ and $r < m$ be given, and consider the following optimization problem

$$\begin{aligned} \underset{x \in X}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & g(x, \delta) \le 0, \text{ for all } \delta \in S \setminus R, \end{aligned} \tag{3}$$

where x, c, X and S are as in (1), and R is a subset of the original sample set with cardinality equal to r. Formulation (3) allows scenarios to be discarded and, as such, can be used to improve the optimal objective value with respect to that of (1). The sampling and discarding approach [7], [8] quantifies the probability of constraint violation associated to the optimal solution (under Assumption 1) of (3), thus providing the decision maker with a sound theoretical result that can be leveraged to trade probability of constraint violation to performance. This result is stated in the sequel.

Theorem *2 (Theorem 2.1, [7]):* Consider Assumption 1, and fix $\epsilon \in (0, 1)$. Let $m > d + r$ and denote by $x^\star(S)$ the optimal solution of (3). If all removed scenarios are violated by the resulting solution $x^\star(S)$, i.e., $g(x^\star(S), \delta) > 0$ for all $\delta \in R$, with $\mathbb{P}^m$-probability one, then

$$\mathbb{P}^m \Big\{ (\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\big\{\delta \in \Delta : g(x^\star(S), \delta) > 0\big\} > \epsilon \Big\} \le \binom{r+d-1}{r} \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{4}$$

As opposed to the tight bound of Theorem 1, the bound of Theorem 2 is not tight. Indeed, in Section 4.2 of [7], the authors show that there exists a class of convex optimization programs and a discarding scheme such that the right-hand side of (4) can be replaced by $\sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$; however, the argument is not constructive and is limited to an existential statement. This raises the following natural question: can one provide tight bounds for a scenario program with discarded constraints? In this paper, we answer the last question by providing a class of optimization problems and a scenario discarding scheme for which this is the case.

C. Learning theoretic background

The main results of this paper are based on some concepts associated to the learning literature and on the interpretation of scenario theory within this framework given in [18].

Definition *3 (Compression set):* Given $m \in \mathbb{N}$. Let $S = (\delta_1, \ldots, \delta_m)$ be an element of Δ^m. Define a mapping $\mathcal{A} : \Delta^m \to 2^\Delta$, where 2^Δ represents the power set of Δ. A subset $C \subset S$ with cardinality equal to ζ is a compression for $\mathcal{A}$ if $\delta \in \mathcal{A}(C)$ for all $\delta \in S$.

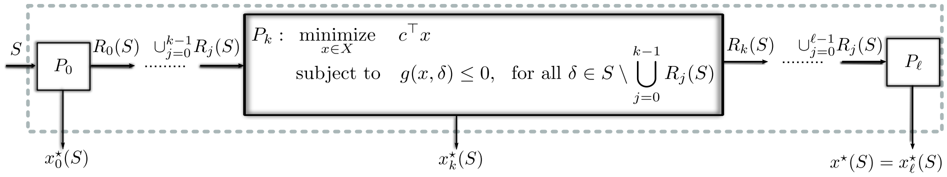

Fig. 1: Block diagram of the proposed scheme. For a given set of scenarios $S = \{\delta_1, \dots, \delta_m\}$ with $(\ell+1)d < m$, we solve a cascade of $\ell+1$ optimization programs denoted by $P_k,\ k \in \{0, \dots, \ell\}$. For each k, $k \in \{0, \dots, \ell-1\}$, we remove $R_k(S)$ scenarios with $|R_k(S)| = d$ as in (6), hence, in total $r = \ell d$ scenarios (the ones in $\bigcup_{j=0}^{\ell-1} R_j(S)$) are discarded. The choice of each set of discarded scenarios depends on the initial set S, thus we introduce it as argument of R_k. The final solution is denoted by $x^\star(S) = x_\ell^\star(S)$.

Roughly speaking, a subset C is a compression for the mapping $\mathcal{A}$ if it assigns the same label to all scenarios in S as if all the samples in S were used. In the statistical learning theory this property is also known as consistency of $\mathcal{A}(C)$ with respect to the samples [17]–[19]. The notion of compression set is crucial to characterize generalization properties of the mapping under unseen scenarios. In fact, after adapting the notation of paper [18] to the one of this paper, we can produce the following result.

***Theorem** 3 (Theorem 3, [18]):* Fix $\epsilon \in (0, 1)$ and $\zeta < m$. If there exists a unique compression set C of cardinality ζ, then

$$\mathbb{P}^m \Big\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\big\{ \delta \in \Delta : \delta \notin \mathcal{A}(C) \big\} > \epsilon \Big\} = \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \quad (5)$$

Note that for a fixed $\epsilon \in (0, 1)$ the right-hand side of (5) tends to zero as m tends to infinity. Hence, Theorem 3 states that is such a unique compression set exists the mapping $\mathcal{A}$ is at least $(1-\epsilon)$-accurate as an approximation of Δ (approximately correct), with confidence (probably) equal to $1 - \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$.

III. The case of fully-supported scenario programs

A. The proposed removal scheme

In this section we present the proposed removal scheme for the sampling and discarding approach. To this end, let $m \in \mathbb{N}$ and assume m i.i.d. samples from the unknown distribution $\mathbb{P}$ are available. Let d be the dimension of the given optimization problem. Fix $r = \ell d$ and, for all $k \in \{1, \dots, \ell\}$, consider

$$P_k: \ \underset{x \in X}{\text{minimize}} \quad c^\top x$$
$$\text{subject to} \quad g(x, \delta) \le 0, \quad \text{for all } \delta \in S \setminus \bigcup_{j=0}^{k-1} R_j(S),$$

where, for $j \in \{0, \dots, \ell-1\}$, $R_j(S)$ represents a subset of S with cardinality d, consisting of the removed scenarios at the j-th stage, and c, X, and $g : \mathbb{R}^d \times \Delta \to \mathbb{R}$ are defined as in (1) and (3). For $k = 0$, we define P_0 as in (1), where all the scenarios are enforced.

Assumption *2:* We assume that, for all $k \in \{0, \dots, \ell\}$,

i) The solution of P_k exists and is unique with $\mathbb{P}^m$ probability one, and its feasible set has non-empty interior;
ii) P_k is fully-supported with $\mathbb{P}^m$ probability one.

For each $k \in \{0, \dots, \ell\}$, we denote by $x_k^\star(S)$ the (unique under Assumption 2) optimal solution of problem P_k. By Assumption 2, item ii), the size of the support set of P_k is equal to d with $\mathbb{P}^m$ probability one. Hence, for all $j \in \{0, \dots, \ell-1\}$, we define

$$R_j(S) = \mathrm{supp}(x_j^\star(S)), \quad (6)$$

which represents the support set at the j-th stage. Even though at the ℓ-th stage of the proposed scheme no scenarios are discarded, we denote by $R_\ell(S)$ the support set associated to the optimal solution of P_ℓ, i.e., we set $R_\ell(S) = \mathrm{supp}(x_\ell^\star(S))$. The final decision under this procedure, also denoted by $x^\star(S)$, is equal to $x_\ell^\star(S)$, the solution for the last optimization problem.

To clarify the structure of the proposed removal scheme, we present in Figure 1 a diagram containing its main components. The scheme has S scenarios as input and, at each stage, the support set of the corresponding problem P_k is removed. To further illustrate how the proposed scheme works, we consider the pictorial example of Figure 2. Note that $d = 2$, $m = 6$, and we remove $r = 4$, thus requiring 3 steps of the removal scheme of Figure 1. All the problems P_k, $k \in \{0, 1, 2\}$, are fully-supported, thus satisfying Assumption 2, item ii). The objective function is given by $c^\top x = x_2$ and is indicated by the downwards pointing arrow. The corresponding solution for the intermediate problem is illustrated by $x_k^\star(S)$, for $k \in \{0, 1, 2\}$, and the support set of each stage by different colour patterns. For instance, the green constraints form $\mathrm{supp}(x_0^\star(S))$, i.e., the support set of P_0. The shaded colour under each constraint corresponds to the region of the plane that violates that given constraint, e.g., we notice that $x_1^\star(S)$ violates both scenarios that belong to $\mathrm{supp}(x_0^\star(S))$ and satisfies all the remaining ones.

***Remark** 1:* Note that, for each $k \in \{0, \dots, \ell\}$, the support scenarios $R_k(S)$ coincide with the active constraints at the optimal solution of P_k due to Assumption 2. Hence, once we have solved problem P_k the support set $R_k(S)$ can be computed without any substantial computational burden.

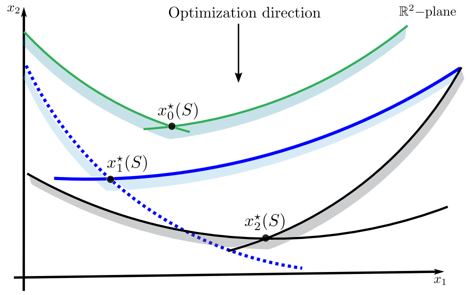

Fig. 2: Pictorial example that illustrates the scheme proposed in Figure 1. In this case, we have that $d = 2,\ m = 6,\ r = 4$, and $\ell = 2$, and all the problems $P_k,\ k \in \{0, 1, 2\}$, satisfy Assumption 2, item ii). The objective function is given by $c^\top x = x_2$ (indicated by the downwards pointing arrow). The support sets are denoted by the different colour patterns. Observe that the dashed-blue constraint is removed by the scheme of Figure 1, but it is not violated by $x_2^\star(S)$.

B. Main result

We are now in the position to present the main result of this paper. Consider the scheme of Figure 1 and let $x^\star(S) = x_\ell^\star(S)$. The next theorem provides guarantees for the probability of constraint violation associated to $x^\star(S)$. Note that the following result is not necessarily tight; in other words, fully-supportedness is not sufficient to guarantee a tight bound on the probability of constraint violation. It needs to be imposed in conjunction with another assumption (see Assumption 3 in the sequel), related to the violation properties of the removed scenarios.

***Theorem** 4:* Consider Assumption 2. Fix $\epsilon \in (0, 1)$, set $r = \ell d$ and let $m > r + d$. Consider also the scenario discarding scheme as encoded by (6) and illustrated in Figure 1, and let the minimizer of the ℓ-th program be $x^\star(S) = x_\ell^\star(S)$. We then have that

$$\mathbb{P}^m\Big\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\big\{\delta \in \Delta : g(x^\star(S), \delta) > 0\big\} > \epsilon\Big\} \leq \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \qquad (7)$$

Before proceeding with the proof of Theorem 4, let us make some comments. Note that the first ℓ iterations of the scheme in Figure 1 can be considered as a particular removal algorithm in the framework of [7]. Being so, we have proved a strictly better bound on the probability of constraint violation that does not involve the combinatorial factor $\binom{r+d-1}{r}$ and that can be achieved by a computationally tractable removal scheme. Besides, for fully-supported problems, Theorem 4 relaxes the assumption in Theorem 2 that requires the optimal solution to violate all the constraints associated to the removed scenarios. For instance, in the pictorial example of Figure 2, the dashed-blue constraint is removed at stage 1, but it is not violated by the final solution of our scheme.

Proof: The proof is divided into three steps. In the interest of space, some of the arguments have been omitted or simplified. We refer the reader to [20] for a detailed proof.

1) Definition of a suitable mapping: Consider Assumption 2. Let $m > (\ell + 1)d$, and consider any set $C \subset S$, with $|C| = (\ell + 1)d$. We consider the proposed scheme of Figure 1, fed by C rather than S. All quantities introduced in this section depending on S will now depend on C instead. For a given set of indices $I \subset C$, we define

$$z^\star(I) = \underset{x \in X}{\operatorname{argmin}} \quad c^\top x \quad \text{subject to} \quad g(x, \delta) \leq 0, \ \text{for all } \delta \in I. \qquad (8)$$

Recall that $x_k^\star(C)$ denotes the minimizer of P_k which in turn is based on the samples in $C \setminus \cup_{j=0}^{k-1} R_j(C)$, i.e., the ones that have not been removed up to stage k of the proposed scheme. It thus holds that $x_k^\star(C) = z^\star(C \setminus \cup_{j=0}^{k-1} R_j(C))$ – note that the argument of $z^\star$ in this case depends on k, $k \in \{0, \dots, \ell\}$. Recall also that $R_k(C) = \operatorname{supp}(x_k^\star(C))$.

Since we will be invoking the framework introduced in Section II-C, we define the mapping $\mathcal{A} : \Delta^m \to 2^\Delta$, with $\zeta = (\ell + 1)d$, as

$$\begin{aligned} \mathcal{A}(C) = \Bigg\{ &\big\{\delta \in \Delta : g(x_\ell^\star(C), \delta) \leq 0\big\} \\ &\cap \Bigg\{ \bigcap_{k=0}^{\ell} \big\{\delta \in \Delta : c^\top z^\star(J \cup \{\delta\}) \leq c^\top x_k^\star(C), \\ &\quad \text{for all } J \subset C \setminus \cup_{j=0}^{k-1} R_j(C), \ \text{with } |J| = d - 1\big\} \Bigg\} \Bigg\} \\ &\cup \Bigg\{ \bigcup_{k=0}^{\ell-1} R_k(C) \Bigg\} = \Big(\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\Big) \cup \mathcal{A}_3(C). \end{aligned} \qquad (9)$$

The main motivation to define the mapping in (9) is the fact that its probability of constraint violation will be shown to upper bound that of $\{\delta \in \Delta : g(x_\ell^\star(C), \delta) \leq 0\}$, which is ultimately the quantity we are interested in.

Note that $\mathcal{A}(C)$ comprises three sets: (1) $\mathcal{A}_1(C)$ contains all realizations of δ for which the final decision of our proposed scheme $x_\ell^\star(C) = x^\star(C)$ remains feasible. This is the set whose probability of occurrence we are ultimately interested to bound; (2) $\mathcal{A}_2(C)$, the intersection of $\ell + 1$ sets, indexed by $k \in \{0, \dots, \ell\}$, each of them containing the realizations of δ such that, for all subsets of cardinality $d - 1$ from the remaining samples at stage k, the cost $c^\top z^\star(J \cup \{\delta\})$ is lower than or equal to $c^\top x_k^\star(C)$. The former cost corresponds to appending δ to any set J of $d - 1$ scenarios from $C \setminus \cup_{j=0}^{k-1} R_j(C)$, while the latter corresponds to the cost of the minimizer $x_k^\star(C)$ of P_k. Informally, this inequality is of similar nature to that of the first set in $\mathcal{A}(C)$, however, rather than considering constraint satisfaction it only involves some cost dominance condition for each of the interim and the final optimal solutions; (3) $\mathcal{A}_3(C)$, which includes all scenarios that are removed by the discarding scheme. Implicit in the definition of mapping (9) is the fact that, for any compression set C, all samples that are not removed in the intermediate stages must be contained in the set $\mathcal{A}_1(C) \cap \mathcal{A}_2(C)$. The following proposition establishes a basic property of any compression associated to the mapping (9).

***Proposition** 1:* Consider Assumption 2. Set $r = \ell d$ and let $m > (\ell+1)d$. We have that $C \subset S$ is a compression set for $\mathcal{A}(C)$ in (9) if and only if, for all $k \in \{0, \ldots, \ell\}$, we have that $x_k^\star(C) = x_k^\star(S)$.

Proof: The proof of this proposition is omitted and can be found in [20]. ■

2) Existence and uniqueness of a compression: A natural compression candidate is

$$C = \bigcup_{k=0}^{\ell} \mathrm{supp}(x_k^\star(S)), \tag{10}$$

as it consists of the support samples for P_k, $k \in \{0, \ldots, \ell\}$.

Existence: Here we only sketch the main arguments. To show that C in (10) is a compression set we show that $x_k^\star(C) = x_k^\star(S)$, $k \in \{0, \ldots, \ell\}$, using induction and then use the sufficiency part of Proposition 1. It is clear that $x_0^\star(C) = x_0^\star(S)$ due to

$$c^\top x_0^\star(S) = c^\top z^\star(S) = c^\top z^\star(\mathrm{supp}(x_0^\star(S))) = c^\top x_0^\star(C),$$

and to Assumption 2, item i). We then assume that $x_j^\star(C) = x_j^\star(S)$ for all $j \in \{0, \ldots, \bar{k}\}$, for some $\bar{k} < \ell$, and show that

$$\begin{aligned} c^\top x_{\bar{k}+1}^\star(C) &= c^\top z^\star(C \setminus \cup_{j=0}^{\bar{k}} R_j(S)) \\ &\le c^\top z^\star(S \setminus \cup_{j=0}^{\bar{k}} R_j(S)) = c^\top x_{\bar{k}+1}^\star(S), \end{aligned} \tag{11}$$

where the first and last equalities are due to (8), and the inequality is due to the fact that $C \setminus \cup_{j=0}^{\bar{k}} R_j(S) \subseteq S \setminus \cup_{j=0}^{\bar{k}} R_j(S)$. Moreover, since $C \subset S$ and $R_j(C) = R_j(S)$ for all $j < \bar{k}$, we have that $c^\top x_{\bar{k}+1}^\star(S) \le c^\top x_{\bar{k}+1}^\star(C)$. This implies that $x_{\bar{k}+1}^\star(C) = x_{\bar{k}+1}^\star(S)$ by Assumption 2, item i), thus concluding the induction proof and showing the existence of a compression set.

Uniqueness: Let C' be another compression set of size $(\ell+1)d$. By Proposition 1 (necessity part), we have that $x_k^\star(C') = x_k^\star(S)$, for all $k \in \{0, \ldots, \ell\}$. Hence, we can conclude that $C = C'$ as these have the same cardinality. This concludes the uniqueness part.

3) Linking Theorem 3 with the probability of constraint violation: Recall that

$$\mathcal{A}(C) = \Big(\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\Big) \cup \mathcal{A}_3(C), \tag{12}$$

where the individual sets are as in (9). Recall also that $\mathcal{A}_3(S)$ is a discrete set. Let $C \subset S$ with $|C| = (\ell+1)d$ be the unique compression defined in (10). We have that

$$\begin{aligned} \mathbb{P}\{\mathcal{A}(C)\} &= \mathbb{P}\{\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\}, \\ &\le \mathbb{P}\{\mathcal{A}_1(C)\} = \mathbb{P}\{\delta \in \Delta : \ g(x^\star(C), \delta) \le 0\}, \\ &= \mathbb{P}\{\delta \in \Delta : \ g(x^\star(S), \delta) \le 0\}, \end{aligned} \tag{13}$$

where the first equality is due to the fact that $\mathbb{P}\{\mathcal{A}_3(C)\} = 0$, since $\mathcal{A}_3(C)$ is a discrete set and $\mathbb{P}$ is non-atomic, which prevents scenarios to have accumulation points with non-zero probability, while the inequality is due to the fact that $\mathcal{A}_1(C) \cap \mathcal{A}_2(C) \subseteq \mathcal{A}_1(C)$. The second to last equality is by definition of $\mathcal{A}_1(C)$, and the last one follows from the fact that $x^\star(C) = x^\star(S)$.

We then have that if $\mathbb{P}\{\delta \in \Delta : \ g(x^\star(S), \delta) > 0\} > \epsilon$ then $\mathbb{P}\{\delta \in \Delta : \ \delta \notin \mathcal{A}(C)\} > \epsilon$. As a result, $\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : \ g(x^\star(S), \delta) > 0\} > \epsilon\} \subseteq \{(\delta_1, \ldots, \delta_m) \in \Delta^m : \ \mathbb{P}\{\delta \in \Delta : \ \delta \notin \mathcal{A}(C)\} > \epsilon\}$. The last statement implies then that

$$\begin{aligned} &\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \ \mathbb{P}\{\delta \in \Delta : \ g(x^\star(S), \delta) > 0\} > \epsilon\} \\ &\le \mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \ \mathbb{P}\{\delta \notin \mathcal{A}(C)\} > \epsilon\}. \end{aligned} \tag{14}$$

Therefore, since the set C in (10) is the unique compression of $\mathcal{A}(C)$, by Theorem 3 we have that

$$\begin{aligned} \mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \ \mathbb{P}\{\delta \in \Delta : \ \delta \notin \mathcal{A}(C)\} > \epsilon\} \\ \le \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \end{aligned} \tag{15}$$

By (14) and (15), we obtain the result. This concludes the proof of Theorem 4. ■

IV. Tightness of the bound of Theorem 4

A. Class of programs for which the bound is tight

The result of this section shows that the bound of Theorem 4 is tight, i.e., there exists a class of convex scenario programs where it holds with equality. To this end, we slightly modify the mapping defined in (9) (see [20] for more details) and, in addition to Assumption 2, we impose the following condition.

Assumption *3:* Fix any $S = \{\delta_1, \ldots, \delta_m\} \in \Delta^m$ and let $C \subset S$. For any $k \in \{0, \ldots, \ell\}$ and $\delta \in S$ such that $\delta \in \mathrm{supp}(x_k^\star(C))$, we have that $g(z^\star(J), \delta) > 0$, for all $J \subset C \setminus \big(\cup_{j=0}^{k-1} \mathrm{supp}(x_j^\star(C)) \cup \{\delta\}\big)$ with $|J| = d$.

Assumption 3 consists in a restriction on the class of fully-supported problems. For instance, the pictorial example of Figure 2 does not satisfy Assumption 3, even though all the intermediate problems P_k are fully-supported, as the dashed-blue removed constraint is not violated by the resulting solution. Indeed, Assumption 3 requires that, with $\mathbb{P}^m$ probability, whenever a sample belongs to the support scenarios of any intermediate problem, then the scenario associated with it is violated by all the solutions that could have been obtained using any subset of cardinality d from the remaining samples. Note that verifying Assumption 4 is hard in general; we show in the next subsection an example that satisfies this requirement and admits an analytic solution. Assumption 3 is similar to the requirement of Theorem 2 [7], [8], however, in Theorem 5 (whose proof is given in [20]) we exploit it in conjunction with the discarding scheme of Figure 1 to show that the result of Theorem 4 is tight.

***Theorem** 5:* Consider Assumptions 2 and 3. Fix $\epsilon \in (0, 1)$, set $r = \ell d$ and let $m > r + d$. Consider also the scenario discarding scheme as encoded by (6) and illustrated in Figure 1, and let the minimizer of the ℓ-th program be $x^\star(S) = x_\ell^\star(S)$. We then have that

$$\begin{aligned} \mathbb{P}^m\Big\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\big\{\delta \in \Delta : g(x^\star(S), \delta) > 0\big\} > \epsilon\Big\} \\ = \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \end{aligned} \tag{16}$$

B. An example with an analytic solution

We revisit the simple problem studied in [7] and show that it satisfies Assumption 3. We use existing results in statistics to compute the violation probability of the solution returned by applying the scheme of Figure 1 to this problem and show that the resulting violation probability coincides with the result of Theorem 5.

To this end, fix $m \in \mathbb{N}$ and $r < m$, and consider the procedure of Section III-A, which involves a sequence of $\ell + 1$ problems. For $k = 0, \dots, r$ (since $d = 1$ in this case), P_k is given by

$$\begin{aligned} &\underset{x \in [0,1]}{\text{minimize}} \quad x \\ &\text{subject to} \quad x \geq \delta_i, \quad i \in S \setminus \cup_{j=0}^{k-1} R_j(S). \end{aligned} \tag{17}$$

We further assume that all samples are extracted from a uniform distribution over the interval $[0, 1]$. Note that (17) satisfies Assumption 2 and Assumption 3 (see [20] for a detailed explanation). Besides, the optimal solution returned by the scheme of Figure 1, due to the fact that $d = 1$, is given by $x^\star(S) = x_r^\star(S) = \max_{i \in S \setminus \cup_{j=0}^{r-1} R_j(S)} \delta_i$, i.e., it is equal to the r-th largest scenario, and the probability of constraint violation is given by $V(x^\star(S)) = 1 - x^\star(S)$ since the distribution $\mathbb{P}$ is uniform. We can then invoke standard results of order statistics [21, Proposition 8.7.1] or the more general result in [11] to conclude that the joint distribution of $(\delta_{(1)}, \dots, \delta_{(m)})$, where $\delta_{(i)}$ represents the i-th largest sample, is a Dirichlet distribution [21, Chapter 7] with all the $(m+1)$ parameters equal to one. As a consequence, the marginal of r-th largest sample $x^\star(S) = \delta_{(r)}$ is a beta distribution with parameters $(m - r + 1, r)$, and the associated probability of constraint violation is given by

$$\mathbb{P}^m\{(\delta_1, \dots, \delta_m) : 1 - x^\star(S) > \epsilon\} = \sum_{i=0}^{r} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i},$$

which agrees with the violation obtained using the result of Theorem 5 for $d = 1$. An alternative derivation based on elementary probability arguments can be found in [20].

V. Conclusion

We revisited the sampling and discarding approach within the scenario approach theory and derived a tight bound on the probability of constraint violation for the obtained solution. To this end, we analyzed a scheme to remove constraints that is composed of a cascade of scenario programs, where at each stage a subset of scenarios related to the support set of the associated optimal solution is removed.

Current work involves showing that the proposed bound is valid, albeit not tight, for any convex scenario program that is not necessarily fully-supported. Under a mild non-degeneracy assumption we have established such a result in [20]. Our present analysis is limited to cases where the number of removed scenarios is a multiple of the dimension of the optimization problem. Current work concentrates towards relaxing this requirement.

VI. Acknowledgement

The authors are grateful to Professors Simone Garatti and Marco Campi for extended discussions on the results presented in this paper. Special thanks to Prof. Garatti for several insightful technical comments.

References

[1] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Automat. Contr.*, vol. 51, no. 5, pp. 742—-753, 2006.

[2] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.

[3] M. Campi and S. Garatti, *Introduction to the Scenario Approach*. MOS-SIAM Series on Optimization, 2018.

[4] M. C. Campi and S. Garatti, "Wait-and-judge scenario optimization," *Math. Program.*, vol. 167, no. 1, pp. 155–189, 2018.

[5] S. Garatti and M. C. Campi, "Complexity-based modulation of the data-set in scenario optimization," in *2019 18th Eur. Control Conf. ECC 2019*, no. 2, 2019, pp. 1386–1391.

[6] G. Calafiore and M. C. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Math. Program.*, vol. 102, no. 1, pp. 25–46, 2005.

[7] M. Campi and S. Garatti, "A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality," *J. Optim. Theory Appl.*, vol. 148, pp. 257–280, 2011.

[8] G. C. Calafiore, "Random convex programs," *SIAM J. Optim.*, vol. 20, no. 6, pp. 3427–3464, 2010.

[9] S. Grammatico, X. Zhang, K. Margellos, P. Goulart, and J. Lygeros, "A Scenario Approach for Non-Convex Control Design," *IEEE Trans. Automat. Contr.*, vol. 61, no. 2, pp. 334–345, 2016.

[10] M. C. Campi, S. Garatti, and F. A. Ramponi, "A General Scenario Theory for Nonconvex Optimization and Decision Making," *IEEE Trans. Automat. Contr.*, vol. 63, no. 12, pp. 4067–4078, 2018.

[11] A. Care, S. Garatti, and M. Campi, "Scenario Min-max Optimization and the Risk of Empirical Costs," *SIAM J. Optim.*, vol. 25, no. 4, pp. 2061–2080, 2015.

[12] G. C. Calafiore, "Repetitive Scenario Design," *IEEE Trans. Automat. Contr.*, vol. 62, no. 3, pp. 1125–1137, 2017.

[13] S. Garatti and M. C. Campi, "Learning for Control: A Bayesian Scenario Approach," in *58th Conf. Decis. Control*. IEEE, 2019, pp. 1772–1777.

[14] D. Paccagnan and M. C. Campi, "The scenario approach meets uncertain variational inequalities and game theory," in *58th Conf. Decis. Control*, Nice, France, 2019, pp. 6124–6129.

[15] F. Fele and K. Margellos, "Probably Approximately Correct Nash Equilibrium Learning," *arXiv:1903.10387 [math.OC]*, pp. 1–12, 2019.

[16] A. Care, S. Garatti, and M. Campi, "A coverage theory for least squares," *J. R. Stat. Soc.*, pp. 1–23, 2018.

[17] M. Vidyasagar, *Learning and Generalization*. Springer, 2002.

[18] K. Margellos, M. Prandini, and J. Lygeros, "On the connection between compression learning and scenario based single-stage and cascading optimization problems," *IEEE Trans. Automat. Contr.*, vol. 60, no. 10, pp. 2716–2721, 2015.

[19] S. Floyd and M. Warmuth, "Sample Compression, Learnability, and the Vapnik-Chervonenkis Dimension," Tech. Rep., 1995.

[20] L. Romao, K. Margellos, and A. Papachristodoulou, "On the exact feasibility of convex scenario programs with discarded constraints," *arXiv:2003.13433 [math.OC]*, pp. 1–15, 2020.

[21] S. Wilks, *Mathematical Statistics*. John Wiley & Sons, Inc., 1962.