Probabilistic feasibility guarantees for convex scenario programs with an arbitrary number of discarded constraints ⋆

Licio Romao, Kostas Margellos, Antonis Papachristodoulou

[a] *Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.*

Abstract

Discarding constraints in scenario optimization, a technique known as the sampling-and-discarding scheme, allows the decision maker to trade feasibility to performance. Recently, a removal scheme with a less conservative bound on the constraint violation probability of the final decision has been proposed. In this letter, we further contribute to the theoretical properties of such a scheme by extending the number of discarded scenarios to be arbitrary, as opposed to an integer multiple of the dimension of the decision space. There are two facets to the results of this paper. On the one hand, our feasibility guarantees outperform the standard "sampling-and-discarding" bound in the literature. On the other hand, we highlight an inherent property of the discarding mechanism, namely, the fact that removing a number of scenarios that is not an integer multiple of the dimension of the decision space is likely to introduce additional conservatism.

Key words: Randomized algorithms, sampling-and-discarding, scenario approach theory, uncertain optimization problems.

1 Introduction

The scenario approach theory consists in a randomized approximation to uncertain optimization problems that involve parameters with a fixed but unknown distribution [5,6,8–10,14,7,11]. At the core of this theory is the so-called scenario program, which consists in an optimization problem whose constraints are enforced based on the available data. Standard results of the scenario approach theory relate feasibility guarantees associated with the optimal solution to the number of available samples and the number of removed scenarios [8,9]. The main theorems in [9,3], which constitute the foundation of the sampling-and-discarding approach to scenario programs, offer feasibility guarantees for any removal scheme and allow the decision maker to trade feasibility to performance. The resulting feasibility bound, however, is not tight, in contrast with a previous result of the scenario approach theory [8] regarding scenario programs without discarded scenarios whose feasibility guarantees hold with equality for the class of the fully-supported scenario programs [8,3]; a formal definition is provided in the sequel.

Recent contributions [18,16] obtain a less conservative bound on the probability of constraint violation than the one given in [9]. The analysis in [18,16] focuses on a specific removal scheme that discards scenarios in an integer multiple of the dimension of decision space by solving a cascade of scenario programs. Another recent paper [17] provides a first step towards a generalization of this procedure to an arbitrary number of discarded scenarios; however, it imposes an assumption that is hard to verify and may not be satisfied apart from problem classes with a specific structure. In this paper, we remove this assumption and propose a new feasibility bound for fully-supported scenario programs that holds for an arbitrary number of removed constraints. There are two facets to our results. On the one hand, we show that our feasibility guarantees for the resulting solution outperform the standard sampling-and-discarding bound in the literature. On the other hand, we highlight an inherent property of the considered removal scheme, namely, the fact that removing a number of scenarios that is not an integer multiple of the dimension of the decision space is likely to introduce additional conservatism.

Hence, our result suggests that – apart from specific cases which are, however, hard to recognize *a priori* (see the results in [17]) – there is no incentive to remove scenarios whose number does not form an integer multiple of the dimension of the decision space. This result complements [9,3] and the recent developments in [18,17], encompassing all possible cases that could

⋆ 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.

Email addresses: licio.romao@eng.ox.ac.uk (Licio Romao), kostas.margellos@eng.ox.ac.uk (Kostas Margellos), antonis@eng.ox.ac.uk (Antonis Papachristodoulou).

emanate within a sampling-and-discarding regime. Moreover, our results are complementary to the ones in [1], where randomized optimization problems are analyzed using the Vapnik-Chervonenkis (VC) theory. The latter results into bounds of similar nature with respect to the proposed ones, however, depend on the VC-dimension which is in general difficult to compute. It is also worth mentioning that our work is in contrast with papers based on randomized sequential algorithms [2,12,4], which require sampling from the unknown distribution in a sequential fashion, rather than relying on an one-shot sampling scheme as in this paper. Sequential algorithms use additional samples to define an exit condition through a validation procedure, which is then employed to assess the guarantees of the final solution.

This paper is organized as follows. In Section 2, we review the sampling-and-discarding approach to scenario programs. In Section 3, we review the removal scheme of [18,16], while the main results of the paper are presented in Section 4. In Section 4.1, we present a motivating example that illustrates the main ideas of this paper. In Section 4.2, we present the extension of the scheme studied in [18] and state the main theorem of this paper. Section 5 compares the proposed bound with that of [9]. The Appendix contains the proof of the main result of Section 4.

2 Background on the scenario approach theory

Let $S = \{\delta_1, \dots, \delta_m\}$ be a collection of independent and identically distributed (i.i.d.) samples from an unknown distribution. We are interested in characterizing properties associated to the optimal solution of

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

with respect to unseen scenarios. We consider $\mathcal{X} \subset \mathbb{R}^d$, $\delta \in \Delta$, with Δ denoting the uncertainty space, $g(x,\delta) : \mathbb{R}^d \times \Delta \to \mathbb{R}$, and $R(S)$ is a subset of S containing scenarios that may have been removed through a possibly iterative procedure. If $R(S) = \emptyset$ then no scenarios are removed. We assume that Δ is endowed with a σ-algebra and there is an unknown probability distribution $\mathbb{P}$ defined on this σ-algebra. Throughout this paper we impose the following assumption.

Assumption 1 *Assume that:*

a. *The solution of problem* (1) *exists and is unique.*
b. *The set $\mathcal{X}$ is closed and convex, and its interior is non-empty.*
c. *The function $g(\cdot,\delta) : \mathbb{R}^d \to \mathbb{R}$ is convex for any $\delta \in \Delta$ and $g(x,\cdot) : \Delta \to \mathbb{R}$ is measurable for any $x \in \mathbb{R}^d$.*

Problem (1) is called scenario program, as its constraints are enforced based on the available scenarios in $S \setminus R(S)$. As apparent in the notation, the choice of $R(S)$ depends on the samples in S, which then implies that the optimal solution of (1) is a random variable defined on Δ^m; to emphasize this dependency we denote it by $x^\star(S)$. The uncertainty space Δ induces both a natural σ-algebra on Δ^m and a probability measure $\mathbb{P}^m$ due to the i.i.d. assumption on [1] S. Assumption 1 imposes mild restrictions on (1). Existence of the solution is guaranteed, for instance, if we consider set $\mathcal{X}$ to be compact. Uniqueness of the optimal solution can always be guaranteed by means of a tie-break rule, e.g., choosing the optimizer with the smallest norm. Non-emptiness of the interior is a standard assumption present in the main results of the scenario theory [5,6,8,9].

Definition 1 (Violation probability) *The function $V : \mathbb{R}^n \to \mathbb{R}$ defined as*

$$V(x) = \mathbb{P}\{\delta \in \Delta : g(x,\delta) > 0\}.$$

denotes the violation probability associated to x.

We are interested in $V(x^\star(S))$, hereafter called the probability of constraint violation, as it measures the risk of violating the constraints for unseen scenarios, not used to obtain $x^\star(S)$.

The scenario approach theory produces bounds on the tail distribution of $V(x^\star(S))$, as stated in [9,3], given by

$$\begin{aligned} &\mathbb{P}^m\{S \in \Delta^m : V(x^\star(S)) > \epsilon\} \\ &\qquad \le \binom{r+d-1}{r} \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}, \end{aligned} \tag{2}$$

where $r = |R(S)|$ is the number of discarded scenarios. Throughout this paper, we will refer to bounds on the tail distribution of the violation probability as feasibility bounds. Notice that the feasibility bound (2) is valid under the assumption that all discarded scenarios are violated by $x^\star(S)$. Besides, given m, ϵ, and d, the bound in (2) allows the decision maker to trade feasibility to performance by discarding scenarios in (1), as the resulting feasible set is enlarged when r increases. The left-hand side of (2) denotes the probability of constraint violation for the solution $x^\star(S)$, while the fact that we allow for $r \neq 0$, implies that the performance/cost $c^\top x^\star$ can only improve compared to the case where $r = 0$. Therefore, increasing r reduces the cost and the bound in (2) allows to control the probability of constraint violation, thus trading probabilistic feasibility to performance.

Key concepts to obtain (2) include the definition of support constraints, and fully-supported programs.

Definition 2 (Support constraints, [8]) *Consider the scenario program in* (1)*. A scenario in $S \setminus R(S)$ is said to be a support scenario (or support constraint) if*

[1] With a slight abuse of notation, throughout the paper we use S to denote a subset of Δ of cardinality equal to m, writing $S \subset \Delta$, or as an element in the product space Δ^m, writing $S \in \Delta^m$.

its removal results in a change in the optimal solution of (1)*. The set of all support scenarios is called the support set of* (1)*, which will be denoted by* $\mathrm{supp}(x^\star(S))$*.*

Definition 3 (Fully-supported problems, [8]) *A scenario program as in* (1) *is said to be fully-supported if for all* $m \in \mathbb{N}$ *the cardinality of the support set is equal to* d *with probability one with respect to* $\mathbb{P}^m$*.*

The notion of fully-supported scenario programs is at the core of the scenario approach theory, especially due to the fact that [8] proves that inequality (2) holds with equality for such programs when no scenarios are discarded (i.e., whenever $r = 0$).

In this paper we are interested in the case where scenarios can be removed (i.e., $r \neq 0$ in (1)). It was elusive whether there would exist a class of scenario programs for which (2) holds with equality. Only recently papers [18,16] show, by analyzing a specific removal algorithm, that such a problem class exists. One of the concepts at the core of the analysis of [18] is that of compression set.

Definition 4 (Compression set, [13,15]) *Let* S *be a set of i.i.d. scenarios from an unknown probability distribution* $\mathbb{P}$*, with* $|S| = m$*. Consider a mapping* $\mathcal{A} : \Delta^m \to 2^\Delta$*, where* 2^Δ *represents the power set of* Δ*. We say that a subset* C *of* S *with* $|C| = \zeta$ *is a compression set of cardinality equal to* ζ *for the mapping* $\mathcal{A}$ *if for all* $\delta \in S$ *we have that, with* $\mathbb{P}^m$*-probability one,* $\delta \in \mathcal{A}(C)$*, which denotes the output of the mapping* $\mathcal{A}$ *using only the samples in* C *as input.*

Notice the slight abuse of notation, where we use the same symbol $\mathcal{A}$ for the mapping that takes as input the set C, which belongs to Δ^ζ as opposed to Δ^m. One of main results of [8] can be interpreted under the lens of the compression set definition given above.

Theorem 1 (Theorem 3, [15]) *Fix any* $\epsilon \in (0, 1)$*. Suppose that the mapping* $\mathcal{A} : \Delta^m \to 2^\Delta$ *possesses a unique compression set of size* ζ*, which we denote by* C*. We then have that*

$$\mathbb{P}^m\{S \in \Delta^m : \mathbb{P}\{\delta \in \Delta : \delta \notin \mathcal{A}(C)\} > \epsilon\} = \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \quad (3)$$

In [15], in the context of scenario optimization, the mapping $\mathcal{A}$ is constructed as

$$\mathcal{A}(C) = \{\delta \in \Delta : g(x^\star(C), \delta) \leq 0\}, \quad (4)$$

where $x^\star(C)$ is the optimal solution of (1) when $r = 0$ (no scenario removal is considered). In particular, it is shown that the existence of a compression set C is related to the underlying problem being fully supported, and in fact the (unique) compression set coincides with the support set of (1), namely, $\mathrm{supp}(x^\star(S))$. Moreover, the set C that constitutes a compression in this case is such that $x^\star(C) = x^\star(S)$, i.e., solving the problem only with the compression set "compresses" the necessary information, and returns the same solution had all the sampled been employed.

If we substitute into (3) the mapping $\mathcal{A}$ defined in (4) we recover the fact that inequality (2) holds with equality for fully-supported programs when no scenarios are discarded, which is one of the main results of [8]. This witnesses the close connection between the notion of compression sets and the scenario approach theory.

Theorem 1 represents a crucial result towards our developments, as it produces a tight bound for the mapping $\mathcal{A}$ as an approximation of the uncertainty set Δ whenever there exists a unique compression of cardinality ζ. Here, we will exploit that theorem by defining a mapping $\mathcal{A}$ different from (4) to account for the case where scenarios are discarded.

3 Removing scenarios in integer multiples of d

We now review the removal scheme proposed in [18]. Consider the scenario program as in (1) and let r be given. Write $r = q_1 d + q_2$, where q_1 and q_2 are integers and $0 \leq q_2 < d$, using the division algorithm. The algorithm described in this section, and studied in detail in [18], is valid only in the case where $q_2 = 0$. The adaptation of this procedure to include an arbitrary number of removed constraints that is not necessarily an integer multiple of d will be presented in Section 4.

For each $k \in \{0, \dots, q_1\}$, consider a sequence of scenario programs given by

$$\begin{aligned} P_k : \underset{x \in \mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & g(x, \delta) \leq 0, \quad \delta \in S \setminus R_k(S), \end{aligned} \quad (5)$$

where R_0 is the empty set, $R_k(S) = R_{k-1}(S) \cup \mathrm{supp}(x^\star_{k-1}(S))$ contains scenarios that have been removed up to stage k, with $x^\star_k(S)$, $k = \{0, \dots, q_1\}$, representing the optimal solution of problem P_k. This removal procedure results in a cascade of $q_1 + 1$ optimization problems and at each stage the support set of P_k is removed. The final solution of the procedure is given by $x^\star_{q_1}(S)$, and will be denoted by $x^\star(S)$. In other words, $x^\star(S)$ is the optimal solution of a scenario program with $R(S) = R_{q_1}(S)$.

The results of [18] are valid for general non-degenerate scenario programs (see [8] and [18] for more details); however, in this paper we impose the following assumption on (5).

Assumption 2 *For each* $k \in \mathbb{N}$*, the scenario program* P_k *given in* (5) *is fully-supported with* $\mathbb{P}^m$*-probability one.*

In other words, Assumption 2 requires that all scenario programs of the removal procedure are fully-supported. Such an assumption is in general strong and may be difficult to satisfy. We adopt it here to facilitate the presentation of our results, but notice that this could be relaxed while leaving our results unaltered by means of a regularization procedure as given, e.g., in [3,18].

Following the notation employed in [18], we define

$$z^\star(J) := \underset{\substack{x\in\mathcal{X} \\ g(x,\delta)\le 0,\ \delta\in J}}{\operatorname{argmin}} \quad c^\top x, \tag{6}$$

as the optimal solution of a scenario program for an arbitrary subset J of the set of samples in S. Under Assumption 1, this is a single-valued mapping and we have that $x_k^\star(S) = z^\star(S \setminus R_k(S))$. The main result of [18], which is presented below for convenience, establishes that the set of scenarios

$$C = \bigcup_{k=0}^{\ell} \operatorname{supp}(x_k^\star(S)), \tag{7}$$

which contains all the support sets of problems P_k's, $k \in \{0, \dots, q_1\}$, is the unique compression set of a certain mapping, thus yielding a bound similar to that of [9]. The structure of this mapping can be found in the Appendix.

Theorem 2 (Theorem 3, [18]) *Fix $\epsilon \in (0,1)$ and let $r = q_1 d$, $m > r + d$. Under Assumptions 1 and 2, denote by $x^\star(S) = x_{q_1}^\star(S)$ the optimal solution of P_{q_1}. We then have that*

$$\mathbb{P}^m\{S \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^\star(S),\delta) > 0\} > \epsilon\} \le \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{8}$$

Paper [18] also shows that the bound in Theorem 2 is tight, as it holds with equality for a sub-class of fully-supported optimization problems.

4 Removing scenarios arbitrarily

4.1 A motivating example

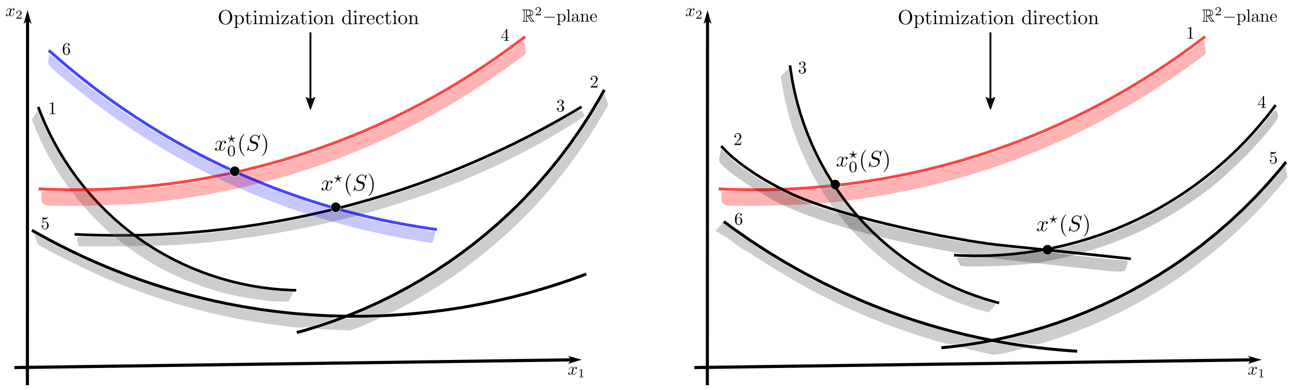
Before presenting the main results of this paper, we introduce two examples that will offer additional interpretation for the subsequent developments. Recall that our ultimate goal is to provide an analysis for a removal strategy that can be used for an arbitrary number of removed scenarios and for all scenario programs. The most natural generalization of the removal procedure described in Section 3 is to proceed as described in the previous section – removing the support scenarios at each stage – and, at the $(q_1 + 1)$-th stage, remove q_2 among the scenarios in $\operatorname{supp}(x_{q_1}^\star(S))$. Under this adaptation, we seek answering the following questions: "To what extent can the analysis carried in [18] be applied to this adapted removal procedure?" and "Does this result in a probability of constraint violation involving a compression set of cardinality equal to $r + d = (q_1 + 1)d + q_2$?"

We start exploring these questions by means of a two-dimensional scenario program with discarded constraints [9,3] when $m = 6$ and $r = 1$ (note that r is not an integer multiple of $d = 2$). Consider a realization illustrated in Figure 1(a), where $x_0^\star(S)$ is the optimal solution of P_0. As $r = 1$, we are not allowed to remove the $\operatorname{supp}(x_0^\star(S))$ as before and need to decide whether to remove the blue or the red scenario in Figure 1(a). In view of obtaining a unique compression set, one cannot allow for such ambiguity; hence, we consider ordered scenarios and associate a label to each constraint in Figure 1(a). Our rule to choose a scenario from $\operatorname{supp}(x_0^\star(S))$ is that of choosing the one with smallest label, which then results in discarding the scenario highlighted in red in Figure 1(a).

Following this rationale, a natural conjecture on the basis of Theorem 2 would be to establish the existence and uniqueness of a compression set of size three. An intuitive candidate is the set composed by the three scenarios supporting both $x_0^\star(S)$ and $x^\star(S)$ in Figure 1(a), as these seem to be sufficient to obtain the same intermediate solutions if only these three scenarios are used. However, in the setting of Figure 1(b) this may not be the case. In fact, following the considered removal procedure, the scenario highlighted in red will be removed in the first iteration of the scheme, thus resulting in the final decision denoted by $x^\star(S)$ in Figure 1(b). Note, however, that differently from the previous realization in Figure 1(a), the support set associated to our final decision does not share scenarios with the support set of the previous stage, hence the individual support sets are disjoint. In fact, any subset of size 3 in the realization of Figure 1(b) would produce distinct interim solutions from $x_0^\star(S)$ and $x^\star(S)$, and this suggests that there is no compression set of size 3 for the realization of Figure 1(b). Such an instance can happen with non-zero probability for distributions that admit a density. Hence, these examples illustrate that for generic cases where the interim support sets do not overlap the compression set cardinality may no longer be $r + d$ as in Theorem 2 but, as we will show in the next section, it is $\lceil r \rceil_d + d$, where $\lceil \cdot \rceil_d$ denotes the smallest integer multiple of d that is greater than r.

4.2 Main result

Consider the removal procedure described in Section 3, and recall that it consists of a cascade of $q_1 + 1$ optimization problems. When $q_2 \neq 0$, we need to remove q_2 out of the d scenarios from $\operatorname{supp}(x_{q_1}^\star(S))$. As motivated in the previous section, we perform such a choice by ordering the samples in S. Formally, this can be done by means of a bijection $\sigma : \{1, ..., m\} \to S$ that assigns an integer from 1 to m to each sample in S. Using such an ordering, for any $\delta_i, \delta_j \in S$, we say that δ_i is smaller than, or equal to, δ_j if $\sigma^{-1}(\delta_i) \le \sigma^{-1}(\delta_j)$ in the usual sense.



(a) Overlapping support sets. (b) Disjoint support sets.

Fig. 1. Two realizations of the scenario program given in (1) with $d = 2$, $m = 6$, and $r = 1$. The ordering of the scenarios is indicated next to each constraint. The solution obtained in the first stage of the process is denoted by $x_0^\star(S)$. The blue and red scenarios correspond to $\mathrm{supp}(x_0^\star(S))$. The red scenario is removed in the first stage of the procedure, as it corresponds to the scenario in $\mathrm{supp}(x_0^\star(S))$ with the smallest label. The final solution is denoted by $x^\star(S)$. In case (a) the support sets for $x_0^\star(S)$ and $x^\star(S)$ overlap, while in case (b) they are disjoint.

Strict inequalities can be interpreted analogously. The feasibility bounds presented in this paper (see Theorem 3 below) hold for any choice of the bijection; however, the optimal objective value depends on that choice. Investigating this effect is outside the scope of this paper.

We then define the optimal solution of the procedure as $x^\star(S) = z^\star(S \setminus R_{q_1+1}(S))$, where $R_{q_1+1}(S) = R_{q_1}(S) \cup \bar{R}(S)$, with $\bar{R}(S)$ containing the q_2 smallest samples from $\mathrm{supp}(x_{q_1}^\star(S))$. In other words, rather than defining $x^\star(S) = x_{q_1}^\star(S)$, as in [18], we remove q_2 samples from $\mathrm{supp}(x_{q_1}^\star(S))$ by composing a set $\bar{R}(S)$. Then we append $\bar{R}(S)$ to $R_{q_1}(S)$ and solve the resulting scenario program with constraints in $S \setminus R_{q_1+1}(S)$ being enforced. Note that when d divides r, we have q_2 equal to zero and this procedure becomes identical to the one analyzed in [18] and described in Section 3. The description of the procedure described in Section 3 and its adaptation in this section can be summarized by defining

$$x^\star(S) = \begin{cases} x_{q_1}^\star(S), & \text{if } q_2 = 0; \\ x_{q_1+1}^\star(S), & \text{otherwise.} \end{cases} \tag{9}$$

We can extend the analysis of this removal scheme when $q_2 \neq 0$ and obtain the following feasibility bound on the resulting solution.

Theorem 3 *Fix $\epsilon \in (0,1)$ and let $\lceil r \rceil_d$ be the smallest integer multiple of d that is greater than r, and $m \geq \lceil r \rceil_d + d$. Let $x^\star(S)$ be defined as in (9). Under Assumptions 1 and 2, we have that*

$$\mathbb{P}^m\{S \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^\star(S), \delta) > 0\} > \epsilon\} \leq \sum_{i=1}^{\lceil r \rceil_d + d - 1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{10}$$

The proof of Theorem 3 can be found in the Appendix. It is divided into two steps: the first one consists of removing $q_1 d$ scenarios by means of the procedure analyzed in [18] and recalled in Section 3; and the second one by analyzing the solution of a scenario program from which only a subset of the support scenarios is discarded. The bound in Theorem 3 could be made explicit with respect to the number of samples using the procedure outlined in [3].

Theorem 3 generalizes Theorem 2, as the latter is recovered from the former if $r = q_1 d$ for some $q_1 \in \mathbb{N}$. The quantity $\lceil r \rceil_d$ in the right-hand side of (10) introduces an additional level of conservatism and is necessary to account for realizations as the one depicted in Figure 1(b). If such cases occur with zero probability, or in other words with probability one the scenario programs are as in Figure 1(a), we can offer a tighter bound, with the upper limit in the summation being $r + d$. Proposition 4, item b), in the Appendix shows this fact; a sufficient condition for this to be the case is provided in [17], and refers to a subclass of fully-supported scenario programs.

Overall, Theorem 3 suggests that if the number of removed scenarios is not an integer multiple of d, then the result of Theorem 2 is no longer valid and the cardinality of the compression set is $\lceil r \rceil_d + d$. As such, discarding scenarios that are not an integer multiple of d does not offer any advantage as the guarantees on constraint violation would be the same as if $\lceil r \rceil_d + d$ scenarios are removed. However, removing more scenarios tends to improve the cost. Hence, the trade-off between feasibility and performance is better if scenarios are removed in an integer multiple of the dimension of the space. Note, however, that the bound of Theorem 3 leads to a less conservative behavior compared to the state-of-art bound summarized in (2) of the sampling-and-discarding mechanism [9,3]. We show this numerically in the next subsection.

5 Numerical examples

5.1 *Comparison with the bound in [9]*

Both bounds (2) and (10) produce feasibility guarantees on the optimal solution for a scenario program with discarded scenarios. While bound (2) possesses a combinatorial factor that increases its conservatism, the one in Theorem 3 has a factor $\lceil r \rceil_d$ in the summation which also generates some level of conservatism. Our goal is compare these bounds. To this end, fix m, r, d, and β, and determine the minimum value of ϵ (i.e., the minimum probability of constraint violation) so that the right-hand side of both (2) and (10) is equal to β. This then implies that for such values the inequality $V(x^\star(S)) \leq \epsilon$ holds, with confidence at least $1-\beta$.

Fix $m = 200$ and $\beta = 10^{-6}$. In Figure 2 we plot the ratio between the ϵ returned by (2) and (10) for different values of d and r. If this ratio is greater than one, then the probability of violation ϵ based on (10) is strictly lower compared to the one in (2), hence the result of Theorem 3 would be less conservative than the bound in [9]. The number of discarded constraints is shown in the $x-$axis, where different colors represent distinct values of d as illustrated in the legend. The violation returned by (10) is lower than that returned by (2) for the considered cases, even for the the most unfavorable case when $r = 4$ and $d = 120$. We should also notice that for $r = 24$ and $d = 30$ the ϵ returned by (2) is approximately equal to 0.59, while the one returned by Theorem 3 is 0.29.

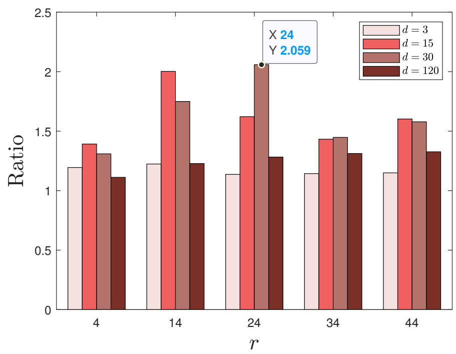


Fig. 2. Comparison between the bounds on the probability of constraint violation for the solution of a scenario program with discarded constraints given in (2) and (10). To obtain these results, we fix $m = 200, \beta = 10^{-6}$ and monitor the ratio between the resulting ϵ from bounds (2) and (10). The $x-$axis shows the number of discarded constraints. Different colors represent distinct values of d.

5.2 *The minimum width interval scenario program*

We now analyze the improvement of the bound of Theorem 3 with respect to inequality (2). We run 10000 runs of a Monte Carlo simulation, where at each run a collection of $m = 200$ i.i.d. samples, denoted by $S = \{\delta_1, \ldots, \delta_m\}$, is generated from a uniform distribution in the interval $[-1, 1]$. We then solve

$$\begin{aligned} \underset{0 \leq x_1 \leq x_2}{\text{minimize}} \quad & x_2 - x_1 \\ \text{subject to} \quad & \delta \in [x_1, x_2], \text{ for all } \delta \in S \setminus R(S), \end{aligned} \qquad (11)$$

where the scenarios in $R(S)$, with $|R(S)| = r = 141$, are removed using the removal scheme described in Section 4. Due to the fact that the distribution is uniform for each Monte Carlo run we obtain an analytic expression for the probability of constraint violation given by

$$V(x^\star(S)) = \frac{2 - (x_2^\star(S) - x_1^\star(S))}{2},$$

i.e., the length of the interval outside $[x_1^\star(S), x_2^\star(S)]$ times the density, which is constant and equal to $\frac{1}{2}$ in this case.

To compare inequality (2) with that of Theorem 3, we construct the empirical cumulative distribution associated with the such a Monte Carlo simulation. In Figure 3 we illustrate the empirical distribution (dashed blue line) and the lower bound on the cumulative distribution given by

$$1 - \sum_{i=0}^{\lceil r \rceil_d + d - 1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i},$$

as dictated by Theorem 3 (solid black line), and

$$1 - \min \left\{ 1, 1 - \binom{r+d-1}{r} \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i} \right\},$$

as in inequality (2) (dash-dotted red line). One can notice that the result of Theorem 3 approximates better the resulting empirical probability distribution, showing the improvement of the proposed bound with respect to the one in (2).

6 Conclusion

In this paper we study fully-supported scenario programs with discarded scenarios by means of a removal scheme that is composed by a cascade of optimization problems. We developed the existing analysis of such a removal procedure to allow for an arbitrary number of removed scenarios. Extensions to deal with non-degenerate scenario programs can be achieved by means of a regularization procedure as in [18,3]. These are not included in this paper for brevity.

An important contribution of this paper is that we generalize the analysis of the removal procedure in [18] to an arbitrary number of removed scenarios. We also highlight an intrinsic limitation of the considered removal scheme, namely, the fact that it is always preferable in terms of achieving a better performance if scenarios are removed in an integer multiple of the dimension of the decision space, and shown that the proposed bound, though not tight, outperforms the one in [9].

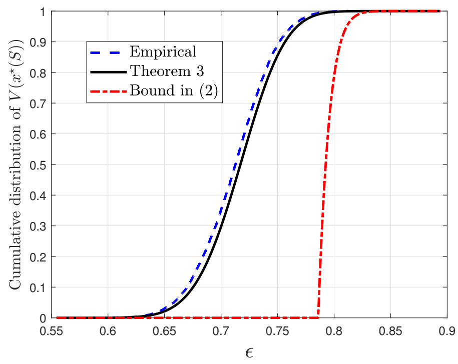


Fig. 3. Comparison between the bound in inequality (2) and of Theorem 3 for $m = 200$, $r = 141$, for the scenario program given by (11). The dashed blue line represents the cumulative distribution of $V(x^\star(S))$ obtained through 10.000 iterations of a Monte Carlo simulation, the solid black line stands for the expression $1-\sum_{i=0}^{\lceil r \rceil_d+d-1}\binom{m}{i}\epsilon^i(1-\epsilon)^{m-i}$ obtained from the the result of Theorem 3, and the dash-dotted red line represents $1-\min\{1, 1-\binom{r+d-1}{r}\sum_{i=0}^{r+d-1}\binom{m}{i}\epsilon^i(1-\epsilon)^{m-i}\}$. Note that the result of Theorem 3 tightly assess the empirical cumulative distribution.

A Appendix: Proof of Theorem 3

The proof of Theorem 3 is divided into two steps. We first study the probability of constraint violation associated to the optimal solution of a scenario program for which only a subset of its support scenarios is removed. Then we combine this analysis with the removal scheme in [18] to produce the bound of Theorem 3.

Step 1: Removing a subset of the support scenarios

Consider a cascade of two scenario programs as in (1) where one is obtained from the other by removing a subset of the support scenarios. Denote these scenario programs by SC_1 and SC_2, respectively, to distinguish them from the P_k in the removal procedure described in Section 3. Let SC_1 be

$$\begin{aligned}\mathrm{SC}_1 : \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & g(x,\delta) \le 0, \quad \delta\in S. \end{aligned} \tag{A.1}$$

Denote by $v^\star(S)$ the optimal solution of (A.1) and denote, as before, by $\mathrm{supp}(v^\star(S))$ its support set. To define SC_2, fix any $0 < q_2 < d$, and let $M(S)$, with $|M(S)| = q_2$, be the subset of $\mathrm{supp}(v^\star(S))$ containing the q_2 smallest scenarios in $\mathrm{supp}(v^\star(S))$ according to an ordering σ (see Section 4.2 for more details). Then, let SC_2 be

$$\begin{aligned}\mathrm{SC}_2 : \underset{x\in\mathcal{X}}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & g(x,\delta) \le 0, \quad \delta\in S\setminus M(S). \end{aligned} \tag{A.2}$$

We denote the optimal solution of (A.2) by $w^\star(S)$ and its support set by $\mathrm{supp}(w^\star(S))$. To analyze the probability of constraint violation properties associated to $w^\star(S)$, we first define, for an arbitrary set of samples $C \subset S$, the set $N(C)$ that contains the smallest scenarios (according to the order defined by σ) that neither support $v^\star(C)$ nor $w^\star(C)$ and that has cardinality equal to that of $\mathrm{supp}(v^\star(C)) \cap \mathrm{supp}(w^\star(C))$. In other words, $N(C)$ contains the $|\mathrm{supp}(v^\star(C)) \cap \mathrm{supp}(w^\star(C))|$-th smallest scenarios of $C \setminus \{\mathrm{supp}(v^\star(C)) \cup \mathrm{supp}(w^\star(C))\}$.

The reader may refer to Figure 1 for a motivation to the definitions of SC_1 and SC_2. In a comparison with the notation of Figure 1 we have that $v^\star(S) = x_0^\star(S)$ and $w^\star(S) = x^\star(S)$ (i.e., SC_1 plays the role of P_0 and SC_2 that of P_1); hence $|\mathrm{supp}(v^\star(C)) \cap \mathrm{supp}(w^\star(C))|$ is equal to the number of scenarios that belong to both support sets of SC_1 and SC_2, e.g., the scenarios are depicted in red in Figure 1. To encompass the fact that the realization in Figure 1(b) may happen with non-zero probability and to obtain a compression set with a cardinality that is uniform with respect to possible realizations, we need to append additional scenarios by forming the set $N(C)$ above.

Similarly as in the proof of Theorem 2, we establish a guarantee on the probability of constraint violation associated to $w^\star(S)$ by showing that there exists a compression scheme associated with such a removal procedure. To this end, we introduce the mapping $\mathcal{B} : \Delta^m \to 2^\Delta$

$$\begin{aligned}\mathcal{B}(C) = \{&\mathcal{B}_1(C)\cap\mathcal{B}_2(C)\cap\mathcal{B}_3(C)\} \\ &\cup \bigcup_{\delta\in M(C)\cup N(C)} \delta, \end{aligned} \tag{A.3}$$

with $\mathcal{B}_1(C) = \{\delta\in\Delta : g(v^\star(C),\delta) \le 0\}$, $\mathcal{B}_2(C) = \{\delta\in\Delta : g(w^\star(C),\delta)\le 0\}$, and

$$\mathcal{B}_3(C) = \Big\{\delta\in\Delta : \delta \ge_\sigma \max_{\xi\in N(C)} \xi\Big\} \cup \mathrm{supp}(w^\star(C)).$$

The set $\mathcal{B}_1(C)\cap\mathcal{B}_2(C)$ contains the scenarios that satisfy both of the interim solutions $v^\star(C)$ and $w^\star(C)$, while $\mathcal{B}_3(C)$ contains scenarios that are either larger than or equal to the maximum scenario [2] in $N(C)$ or that are in $\mathrm{supp}(w^\star(S))$. In fact, the next proposition shows that

$$C = \mathrm{supp}(v^\star(S)) \cup \mathrm{supp}(w^\star(S)) \cup \bigcup_{\delta\in N(S)} \delta \tag{A.4}$$

is the unique compression set for (A.3).

Proposition 4 *Let $0 < q_2 < d$ be a given integer. Consider the cascade of two scenarios programs SC_1 and SC_2 as in (A.1) and (A.2), respectively. The following statements hold:*

[2] Formally, the ordering σ^{-1} is only defined on the finite set S. However, given any finite set S and under mild conditions on the uncertainty space Δ, one may extend σ^{-1} to the whole space Δ in a way that its restriction to S is the original bijection.

a) Suppose that the realization of Figure 1(b) happens with non-zero probability, i.e., suppose that, for all $m \in \mathbb{N}$, $\mathbb{P}^m\{S \in \Delta^m : |\mathrm{supp}(v^\star(S)) \cap \mathrm{supp}(w^\star(S))| = 0\} > 0$. Then, we have that:

1) There exists a realization of scenarios S such that no compression of size smaller than $2d$ exists for the mapping $\mathcal{B}$ in (A.3).

2) The set C in (A.4) is the unique compression set of cardinality $2d$ for the mapping $\mathcal{B}$ in (A.3).

b) If the realization depicted in Figure 1(b) happens with probability zero, i.e., if for all $m \in \mathbb{N}$ we have that $\mathbb{P}^m\{S \in \Delta : |\mathrm{supp}(v^\star(S)) \cap \mathrm{supp}(w^\star(S))| = 0\} = 0$, then there exists a unique compression set of cardinality equal to $q_2 + d$.

Remark 1 *Proposition 4 establishes compression properties related to a removal scheme that discards only a subset of the support scenarios of a scenario program, i.e., the set $M(C)$ above. A striking feature of this scheme is the fact that in the general case (item a)) it may not yield tight bounds on the probability of constraint violation associated to $w^\star(C)$, as we may not have a compression set of cardinality equal to $d + q_2 < 2d$.*

PROOF. *Item a.*1). We argue by contradiction. Let $S \subset \Delta$ be a set with cardinality m and assume that there exists a compression C' of cardinality $d' < 2d$ for the mapping $\mathcal{B}$ in (A.3). Fix a realization S that yields $N(S) = \emptyset$, i.e., one in which the support sets $\mathrm{supp}(v^\star(S))$ and $\mathrm{supp}(w^\star(S))$ are disjoint (e.g., see Figure 1(b)). Note that such a realization exists due to the assumption of item a). As the cardinality of C' is strictly smaller than $2d$ we can find a scenario in $\{\mathrm{supp}(v^\star(S)) \cup \mathrm{supp}(w^\star(S))\} \setminus C'$, since the union of the support sets has cardinality equal to $2d$.

Let $\bar{\delta}$ be an element in $\{\mathrm{supp}(v^\star(S)) \cup \mathrm{supp}(w^\star(S))\} \setminus C'$. Such a $\bar{\delta}$ is either in $\mathrm{supp}(v^\star(S)) \setminus C'$ or in $\mathrm{supp}(w^\star(S)) \setminus C'$. Assume that $\bar{\delta} \in \mathrm{supp}(v^\star(S)) \setminus C'$, then the set $\mathrm{supp}(v^\star(S)) \setminus C'$ is non-empty. We next show that there exists a $\bar{\delta} \in \mathrm{supp}(v^\star(S)) \setminus C'$ such that $g(v^\star(C'), \bar{\delta}) > 0$. Recall that by the definition of a compression set we must have $g(v^\star(C'), \delta) \leq 0$ for all $\delta \in S$, so the existence of such a $\bar{\delta}$ implies that $\mathrm{supp}(v^\star(S))$ must be contained in C'. To this end, suppose for the sake of contradiction that $g(v^\star(C'), \bar{\delta}) \leq 0$ for all $\bar{\delta} \in \mathrm{supp}(v^\star(S)) \setminus C'$. This means that $v^\star(C')$ can be obtained by the following scenario program

$$\begin{aligned} &\underset{x \in \mathbb{R}^d}{\text{minimize}} && c^\top x \\ &\text{subject to} && g(x, \delta) \leq 0, \quad \delta \in C' \cup \mathrm{supp}(v^\star(S)), \end{aligned}$$

as adding the scenarios in $\mathrm{supp}(v^\star(S)) \setminus C'$ does not change the optimal cost. However, by the definition of support set and due to Assumption 1, this implies that $v^\star(C') = v^\star(S)$, which contradicts the fact that $\mathrm{supp}(v^\star(S)) \setminus C'$ is non-empty. Hence, we must have $g(v^\star(C'), \bar{\delta}) > 0$; however, this contradicts the fact that C' is a compression set for the mapping $\mathcal{B}$ in (A.3). In other words, if C' is a compression set of cardinality d then $\bar{\delta} \in \mathrm{supp}(w^\star(S)) \setminus C'$.

Since $\mathrm{supp}(v^\star(S)) \subset C'$, we must have that $v^\star(S) = v^\star(C')$ by Assumption 1, which then implies $M(S) = M(C')$. Changing S by $S \setminus \{\mathrm{supp}(v^\star(S)) \cup M(S)\}$ and C' by $C' \setminus \{\mathrm{supp}(v^\star(S)) \cup M(S)\}$ we can argue similarly as above to conclude that if $\mathrm{supp}(w^\star(S)) \setminus C'$ is not empty, then we can find an element in $\bar{\delta} \in \mathrm{supp}(w^\star(S)) \setminus C'$ such that $g(w^\star(C'), \bar{\delta}) > 0$, which contradicts the fact that C' is a compression. This concludes the proof of item a.1).

*Item a.*2). *(Existence)* We start the proof by showing that the set (A.4) is a compression for the mapping $\mathcal{B}$ in (A.3). To this end, we need to show that $\delta \in \mathcal{B}(C)$ for all $\delta \in S$. By the choice of C in (A.4) and under Assumption 1, we note that $v^\star(C) = v^\star(S)$ and $w^\star(C) = w^\star(S)$, which then implies $M(C) = M(S)$ and $N(C) = N(S)$. Pick $\bar{\delta} \in C$ and let us show that $\bar{\delta} \in \mathcal{B}(C)$. Suppose $\bar{\delta} \in \mathrm{supp}(v^\star(C))$. In this case we have two options: (1) either $\bar{\delta} \in M(S)$, which belongs to the discrete part of $\mathcal{B}(C)$; or (2) $\bar{\delta} \notin M(S)$, in which case it can be either in the support of $\mathrm{supp}(w^\star(S))$ or not. If $\bar{\delta} \in \mathrm{supp}(w^\star(S))$, then it belongs to $\mathcal{B}_1(C) \cap \mathcal{B}_2(C) \cap \mathcal{B}_3(C)$. The fact that such a $\bar{\delta}$ belongs to $\mathcal{B}_1(C) \cap \mathcal{B}_2(C)$ is clear due to $g(v^\star(S), \bar{\delta}) \leq 0$ and $g(w^\star(S), \bar{\delta}) \leq 0$, while $\bar{\delta} \in \mathcal{B}_3(C)$ follows by definition, since $\mathrm{supp}(w^\star(S)) \subset \mathcal{B}_3(C)$. Otherwise, if $\bar{\delta} \in \mathrm{supp}(v^\star(S)) \setminus \mathrm{supp}(w^\star(S))$ then it either belongs to $N(S)$, which then implies that $\bar{\delta} \in \mathcal{B}(C)$, or $\bar{\delta} \in \mathrm{supp}(v^\star(S)) \setminus \{\mathrm{supp}(w^\star(S)) \cup N(S)\}$, hence it belongs to $\mathcal{B}_1(C) \cap \mathcal{B}_2(C)$ by definition, and to $\mathcal{B}_3(C)$ due to the fact that such a $\bar{\delta}$ must satisfy $\bar{\delta} \geq_\sigma \max_{\xi \in N(S)} \xi$. This shows that $\delta \in \mathcal{B}(C)$ for all $\delta \in \mathrm{supp}(v^\star(C))$.

Suppose now that $\bar{\delta} \in \mathrm{supp}(w^\star(C))$. It is straightforward to show that $\bar{\delta} \in \mathcal{B}_1(C) \cap \mathcal{B}_2(C) \cap \mathcal{B}_3(C)$ by means of similar arguments as above, so we have that $\bar{\delta} \in \mathcal{B}(C)$. Besides, if $\bar{\delta} \in N(C)$, then it belongs to the discrete part of $\mathcal{B}(C)$. Therefore, in any case if $\bar{\delta} \in C$, then $\bar{\delta} \in \mathcal{B}(C)$.

To conclude the existence proof, we need to show that if $\bar{\delta} \in S \setminus C$ then $\bar{\delta} \in \mathcal{B}(C)$. Since such a $\bar{\delta}$ is not in the discrete part of the mapping $\mathcal{B}(C)$, we need to show that $\bar{\delta} \in \mathcal{B}_1(C) \cap \mathcal{B}_2(C) \cap \mathcal{B}_3(C)$. As this $\bar{\delta}$ is feasible for both scenarios programs SC_1 and SC_2 we have that $\bar{\delta} \in \mathcal{B}_1(C) \cap \mathcal{B}_2(C)$. It remains to show that $\bar{\delta} \in \mathcal{B}_3(C)$. To this end, note that since $\bar{\delta} \notin C$ we have immediately that $\bar{\delta} >_\sigma \max_{\xi \in N(S)} \xi$, so it belongs to $\mathcal{B}_3(C)$. This shows that C given in (A.4) is a compression set for the mapping $\mathcal{B}$ in (A.3), thus concluding the existence part of the proof.

(Uniqueness) We divide the uniqueness proof into two cases: $N(S) = \emptyset$ and $N(S) \neq \emptyset$. In the former case, let C' be another compression set of size $2d$. Fix any $\bar{\delta} \in C \setminus C'$ and note that either $\bar{\delta} \in \mathrm{supp}(v^\star(C))$ or $\bar{\delta} \in \mathrm{supp}(w^\star(C))$ (note that $\bar{\delta}$ cannot belong to both sets due to the fact that $N(S) = N(C) = \emptyset$ is empty).

If $\bar{\delta} \in \mathrm{supp}(v^\star(S))$ then a similar argument as in item a) (changing S by C in that argument) shows that there exists a $\bar{\delta} \in C \setminus C'$ such that $g(v^\star(C'), \bar{\delta}) > 0$, which contradicts the fact that C' is a compression. A similar argument also holds for $\bar{\delta} \in \mathrm{supp}(w^\star(C))$.

Consider now the case where $N(S) \neq \emptyset$. We proceed similarly as to the previous case and let C' be another compression of size $2d$. Fix any $\bar{\delta} \in C \setminus C'$ and note that $\bar{\delta}$ cannot belong to $\mathrm{supp}(v^\star(C)) \cup \mathrm{supp}(w^\star(C))$, as this would contradict, as before, the fact that C' is a compression. Hence, such a $\bar{\delta}$ must be an element of $N(C) \setminus C'$. Besides, since $\bar{\delta} \notin C'$ and C' is a compression, we must have that $\bar{\delta}$ is in $\mathcal{B}_1(C') \cap \mathcal{B}_2(C') \cap \mathcal{B}_3(C')$. However, $\bar{\delta} \notin \mathcal{B}_3(C')$ as we have $\bar{\delta} <_\sigma \max_{\xi \in N(C')} \xi$, due to the fact that $C' \subset S$ and $\bar{\delta} \notin \mathrm{supp}(w^\star(C')) \subset C'$, which imply that

$$\max_{\xi \in N(C')} \xi > \max_{\xi \in N(C) = N(S)} \xi,$$

This contradicts the fact that C' is a compression, thus concluding the proof of item $a.2$).

Item b). The proof of this item is omitted for brevity and can be found in [17]. In fact, note that Proposition 1 of [17] shows that a particular sub-class of fully-supported scenario programs, namely, the one satisfying Assumption 2 in [17], has the property that $\mathbb{P}^m\{S \in \Delta : |\mathrm{supp}(v^\star(S)) \cap \mathrm{supp}(w^\star(S))| = 0\} = 0$ for all $m \in \mathbb{N}$. This is then exploited in Proposition 2 of [17] to prove item b) of Proposition 4.

Step 2: Combining Proposition 4 with [18]

To account for the general case we consider the setting of Proposition 4, item a). We are now in position to prove Theorem 3. Recall that d is the dimension of the optimization problem P_k and we are writing $r = q_1 d + q_2$, with $0 < q_2 < d$, where $m > \lceil r \rceil_d + d$. Define the mapping $\bar{\mathcal{A}} : \Delta^m \to 2^\Delta$ such that

$$\bar{\mathcal{A}}(C) = \mathcal{A}(C) \cap \{\mathcal{B}(C \setminus R_{q_1}(C)) \cup R_{q_1}(C)\}, \quad \text{(A.5)}$$

where $\mathcal{A}$ is the mapping given by

$$\mathcal{A}(C) = (\mathcal{A}_1(C) \cap \mathcal{A}_2(C)) \cup \mathcal{A}_3(C), \quad \text{(A.6)}$$

with, $\mathcal{A}_1(C) = \{\delta \in \Delta : g(x^\star_{q_1}(S), \delta) \leq 0\}$, $\mathcal{A}_3(C) = \bigcup_{k=0}^{q_1 - 1} \mathrm{supp}(x^\star_k(C))$, and

$$\mathcal{A}_2(C) = \left\{ \bigcap_{k=0}^{q_1-1} \left\{ \delta \in \Delta : c^\top z^\star(J \cup \{\delta\}) \leq c^\top x^\star_k(S), \text{ for all } J \subset S \setminus R_k(S), \text{ with } |J| = d - 1 \right\} \right\}. \quad \text{(A.7)}$$

The mapping $\mathcal{A}$ is associated with the removal procedure encoded by (5) when $q_2 = 0$ and has been introduced in [18,16], and $\mathcal{B}$ is the mapping of Proposition 4, item a), with input given by $S \setminus R_{q_1}(S)$, rather than S. Note also that under this choice for the input of $\mathcal{B}$ we have $v^\star(S \setminus R_{q_1}(S)) = x^\star_{q_1}(S)$ and $w^\star(S \setminus R_{q_1}(S)) = x^\star_{q_1+1}(S) = x^\star(S)$ (see Section 4.2). In fact, under this notation, the scenario programs SC_1 and SC_2 in Proposition 4, item a), correspond to P_{q_1} and P_{q_1+1}, respectively, in the description of Section 3.

We will show that the subset of the scenarios given by

$$C = \bigcup_{k=0}^{q_1} \mathrm{supp}(x^\star_k(S)) \cup \mathrm{supp}(x^\star(S)) \cup \bigcup_{\delta \in N(S)} \delta \quad \text{(A.8)}$$

is a compression set for the mapping $\bar{\mathcal{A}}$ in (A.5) – uniqueness will be shown in the sequel. First, note that such a C can be written as

$$C = C_1 \cup C_2, \; C_1 = \bigcup_{k=0}^{q_1} \mathrm{supp}(x^\star_k(S)),$$
$$C_2 = \mathrm{supp}(x^\star_{q_1}(S)) \cup \mathrm{supp}(x^\star(S)) \cup \bigcup_{\delta \in N(S)} \delta. \quad \text{(A.9)}$$

The fact that C in (A.8) forms a compression set for the mapping $\bar{\mathcal{A}}$ follows trivially since C_1 and C_2 are compression sets for the removal procedure encoded by (5) due to Theorem 4 in [18] and Proposition 4, item a), i.e., $\delta \in \mathcal{A}(C) \cap \{\mathcal{B}(C \setminus R_{q_1}(C)) \cup R_{q_1}(C)\}$ for all $\delta \in S$. Besides, observe that the cardinality of C is equal to $(q_1 + 2)d = \lceil q_1 d + q_2 \rceil_d + d = \lceil r \rceil_d + d$ due to definition of set $N(S)$ given in Proposition 4, item a), and to the relation $r = q_1 d + q_2$.

We now show that the set C in (A.8) is the unique compression set of cardinality equal to $\lceil r \rceil_d + d$ for the mapping in (A.5). Suppose C' is another compression set of cardinality equal to $\lceil r \rceil_d + d$ for $\bar{\mathcal{A}}$. This means that $\delta \in \bar{\mathcal{A}}(C')$ for all $\delta \in S$. However, by the results in [18], we must have $C_1 \subset C'$; otherwise, there would exist another compression set of size $(q_1 + 1)d$ for the mapping $\mathcal{A}$. We also obtain that $\delta \in \mathcal{B}(C')$ for all $\delta \in S$. Since $C' \setminus R_{q_1}(S) \subset S \setminus R_{q_1}(S)$, by Proposition 4, we must also have that $C_2 \subset C$. However, as the cardinality of $C_1 \cup C_2$ is equal to $\lceil r \rceil_d + d$, this implies that $C' = C$, thus showing uniqueness of the compression set C in (A.8).

It remains to show how the existence and uniqueness of a compression set for the mapping $\bar{\mathcal{A}}$ can be used to produce the bound of Theorem 3. To this end, recall that (the dependence on C of the inner sets is omitted to simplify the notation)

$$\bar{\mathcal{A}}(C) = \underbrace{\{(\mathcal{A}_1 \cap \mathcal{A}_2) \cup \mathcal{A}_3\}}_{\mathcal{A}(C)} \cap \underbrace{\{(\mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_3) \cup \mathcal{B}_4\}}_{\mathcal{B}(C \setminus R_{q_1}(C)) \cup R_{q_1}(C)},$$

where we have defined $\mathcal{B}_4 = R_{q_1} \cup \bigcup_{\delta \in M \cup N} \delta$, which contains all the removed scenarios and potentially additional scenarios that compose the set $N(C)$ described in Proposition 4. After some manipulations, we show that

$$\begin{aligned}\bar{\mathcal{A}}(C) &\subset (\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_3) \cup (\mathcal{A}_3 \cup \mathcal{B}_4) \\ &= (\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{B}_2 \cap \mathcal{B}_3) \cup (\mathcal{A}_3 \cup \mathcal{B}_4), \qquad \text{(A.10)}\end{aligned}$$

where the second equality holds due to the fact that $x^\star_{q_1}(C) = v^\star(C \setminus R_{q_1}(C))$, which in turn implies that $\mathcal{A}_1(C) = \mathcal{B}_1(C \setminus R_{q_1}(C))$. Our ultimate goal is to bound the probability of $\mathcal{B}_2$. We can then use (A.10) to obtain

$$\begin{aligned}&\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \notin \mathcal{B}_2(C \setminus R_{q_1}(C))\} > \epsilon\} \\ &\qquad \leq \mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \notin \bar{\mathcal{A}}(C)\} > \epsilon\}.\end{aligned}$$

However, note that the left-hand side of the above inequality is the probability of constraint violation we are interested in and the right-hand side can be upper bounded – due to Theorem 1 (or Theorem 3 in [15]) and to the fact that there exists a unique compression set of size $\lceil r \rceil_d + d$ (as shown above) – by the right-hand side of inequality (10). This concludes the proof of Theorem 3.

References

[1] Teodoro Alamo, Roberto Tempo, and Eduardo F. Camacho. Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems. *IEEE Transactions on Automatic Control*, 54(11):2545–2559, 2009.

[2] Teodoro Alamo, Roberto Tempo, Amalia Luque, and Daniel R. Ramirez. Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms. *Automatica*, 52:160–172, 2015.

[3] Giuseppe Calafiore. Random convex programs. *SIAM Journal on Optimization*, 20(6):3427–3464, 2010.

[4] Giuseppe Calafiore. Repetitive Scenario Design. *IEEE Transactions on Automatic Control*, 62(3):1125–1137, 2017.

[5] Giuseppe Calafiore and Marco Campi. Uncertain convex programs: Randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[6] Giuseppe Calafiore and Marco Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742—-753, 2006.

[7] Marco Campi and Algo Carè. Random convex programs with L1-regularization: Sparsity and generalization. *SIAM Journal on Control and Optimization*, 51(5):3532–3557, sep 2013.

[8] Marco Campi and Simone Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.

[9] Marco Campi and Simone Garatti. A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality. *Journal of Optimization Theory and Applications*, 148:257–280, 2011.

[10] Marco Campi and Simone Garatti. *Introduction to the Scenario Approach.* MOS-SIAM Series on Optimization, 2018.

[11] Aldo Carè, Simone Garatti, and Marco Campi. Scenario Min-max Optimization and the Risk of Empirical Costs. *SIAM Journal on Optimization*, 25(4):2061–2080, 2015.

[12] Mohammadreza Chamanbaz, Fabrizio Dabbene, Roberto Tempo, Venkatakrishnan Venkataramanan, and Qing Guo Wang. Sequential Randomized Algorithms for Convex Optimization in the Presence of Uncertainty. *IEEE Transactions on Automatic Control*, 61(9):2565–2571, 2016.

[13] Sally Floyd and Manfred Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(21):269–304, 1995.

[14] Simone Garatti, Marco Campi, and Algo Carè. On a class of interval predictor models with universal reliability. *Automatica*, 110, 2019.

[15] Kostas Margellos, Maria Prandini, and John Lygeros. On the connection between compression learning and scenario based single-stage and cascading optimization problems. *IEEE Transactions on Automatic Control*, 60(10):2716–2721, 2015.

[16] Licio Romao, Kostas Margellos, and Antonis Papachristodoulou. Tight generalization guarantees for the sampling and discarding approach to scenario optimization. In *59th IEEE Conference on Decision and Control*, pages 2228–2233, 2020.

[17] Licio Romao, Kostas Margellos, and Antonis Papachristodoulou. Tight sampling and discarding bounds for scenario programs with an arbitrary number of removed samples. In *3rd Annual Learning for Dynamics & Control Conference.*, Zurich, Switzerland., 2021.

[18] Licio Romao, Kostas Margellos, and Antonis Papachristodoulou. On the exact feasibility of convex scenario programs with discarded constraints. *IEEE Transactions on Automatic Control*, To appear, 2022.