Proceedings of Machine Learning Research vol 144:1–12, 2021

Tight sampling and discarding bounds for scenario programs with an arbitrary number of removed samples

Licio Romao LICIO.ROMAO@ENG.OX.AC.UK
Kostas Margellos KOSTAS.MARGELLOS@ENG.OX.AC.UK
Antonis Papachristodoulou ANTONIS@ENG.OX.AC.UK
Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ, Oxford, United Kingdom

Abstract

The so-called scenario approach offers an efficient framework to address uncertain optimisation problems with uncertainty represented by means of scenarios. The sampling-and-discarding approach within the scenario approach literature allows the decision maker to trade feasibility to performance. We focus on a removal scheme composed by a cascade of scenario programs that removes at each stage a superset of the support set associated to the optimal solution of each of these programs. This particular removal scheme yields a scenario solution with tight guarantees on the probability of constraint violation; however, existing analysis restricts the number of discarded scenarios to be a multiple of the dimension of the optimisation problem. Motivated by this fact, this paper presents pathways to extend the theoretical analysis of this removal scheme. We first provide an extension for a restricted class of scenarios programs for which tight bounds can be obtained, and then we provide a conservative bound on the probability of constraint violation that is valid for any scenario program and an arbitrary number of removed scenarios, which is, however, not tight.

1. Introduction

Data abound in modern applications, and this can be leveraged to boost robustness against uncertainty. In the past decades, new research directions have sprung from this fact, and are now shaping the theoretical foundation of several disciplines, including control theory and machine learning. Under this scenario, several data-driven algorithms came to prominence. In this paper we focus on a specific randomised technique, called the scenario approach theory (Calafiore and Campi (2005, 2006); Campi and Garatti (2008, 2011)), to study uncertain optimisation problems that provides guarantees on the probability of constraint violation associated to the optimal solution of an optimisation problem, called scenario program, whose constraints are enforced on available data.

The main result of Campi and Garatti (2008) consists of a distribution-free upper bound for the probability of constraint violation associated to the optimal solution of a scenario program. This bound involves the number of samples, dimension of the underlying optimisation problem, and desired level of violation and can be exploited by the decision maker to decide the number of samples required to produce a solution with prescribed feasibility properties. The scenario approach theory then enables us to assess the risk of our decision by means of data, thus consisting of a data-driven approach to uncertain optimisation. Another relevant interpretation of the results of Campi and Garatti (2008) is their connection with the well-known chance-constrained formulation (Prékopa (2003); Nemirovski and Shapiro (2006); Pagnoncelli et al. (2009)) of optimisation problems. The scenario approach theory can be interpreted as generating, with high probability, a feasible solution to chance-constrained problems which is, in general, not optimal. Several papers have studied the optimality gap between the chance-constrained formulation and its scenario approximation (Calafiore and Campi (2006); Campi and Garatti (2011); Esfahani et al. (2015)).

Notwithstanding the advances due to the results in Campi and Garatti (2008), and as price for their generality, the performance of a scenario program in terms of cost may be conservative. To this end, paper Campi and Garatti (2011) characterised the probability of violation of a scenario program in which the decision maker is allowed to discard some of the original scenarios. As the feasible set of a scenario program is enlarged when constraints are discarded, the results in Campi and Garatti (2011), known as the sampling-and-discarding approach, enable the decision maker to reduce the conservatism of a scenario solution, while keeping the probability of constraint violation under control. An interesting feature of the bound proposed in Campi and Garatti (2011) is the fact that it holds true for any removal scheme and, similarly to the bound in Campi and Garatti (2008), is distribution-free. However, in contrast with Campi and Garatti (2008), which holds with equality for fully-supported scenario programs, the bound in Campi and Garatti (2011) is not tight.

Recently, the authors in Romao et al. (2020a) and Romao et al. (2020b) studied a specific removal scheme and provided a bound on the probability of constraint violation that outperforms the one in Campi and Garatti (2011). These papers show tightness of the proposed bound by providing a class of scenario programs that achieves this bound with equality. Such a removal scheme is composed by a cascade of scenario programs where at each stage a superset of the support scenarios associated to the optimal solution is removed. However, their analysis restricts the number of discarded scenarios to be an integer multiple of the dimension of decision space.

The main contribution of this paper is to explore the extent to which the analysis developed in Romao et al. (2020a) can be extended to allow for arbitrary discarded scenarios. First, we characterise the class of scenario programs that permits such arbitrary removal. We show that this coincides with the class of problems that led to tight bounds in Romao et al. (2020a). For general scenario programs, we argue that no tight results can be obtained without exploring additional structure of the problem and provide an upper bound on the probability of constraint violation that improves upon the bound in Campi and Garatti (2011). Moreover, if we are dealing with a min-max scenario programs, we also present another alternative to remove scenarios that combines our removal procedure with the strategy presented in Care et al. (2015) and Garatti et al. (2019).

This paper is organised as follows. The main concepts of the scenario approach theory and the removal strategy analysed in Romao et al. (2020a) are presented in Section 2. The main results of this paper, which include a generalisation of the removal procedure of Romao et al. (2020a) to an arbitrary discarded scenarios for a subclass of scenario programs, are presented in Section 3. Section 4 concludes the paper and provides directions for future work.

2. Problem statement

Let Δ represent the uncertainty space, which is endowed with a σ-algebra $\mathcal{F}$, and suppose that there is an unknown probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ defined over $\mathcal{F}$. The triple $(\Delta, \mathcal{F}, \mathbb{P})$ is called a probability measure[1] space and an element $\delta \in \Delta$ is referred as a scenario. Fix $\epsilon \in (0, 1)$, and suppose that we want to find the optimal solution of the chance-constrained problem

$$\begin{aligned} &\underset{x\in\mathcal{X}}{\text{minimise}} \quad c^\top x \\ &\text{subject to} \quad \mathbb{P}\{\delta : g(x,\delta) > 0\} \le \epsilon, \end{aligned} \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^d$ is a closed, convex set with non-empty interior and function $g : \mathbb{R}^d \times \Delta \to \mathbb{R}$ is measurable[2] in the second argument for each $x \in \mathcal{X}$. The feasible set of (1) comprises of points $x \in$

1. We refer the reader to (Salamon, 2016, Chapter 1) for an introduction to these measure theoretic concepts.
2. For a fixed $x \in \mathcal{X}$, the function $g(x, \cdot) : \Delta \to \mathbb{R}$, where $(\Delta, \mathcal{F}, \mathbb{P})$ is a probability measure space, is said to be measurable if for all Borel sets A of $\mathbb{R}$ we have that $g^{-1}(x, A) \in \mathcal{F}$. This implies that the set $\{\delta \in \Delta : g(x, \delta) > 0\} = g^{-1}(x, A)$, with $A = (0, \infty)$, is an element of $\mathcal{F}$, hence rendering (1) well-defined. A Borel set is an element of the σ-algebra generated by the standard topology of $\mathbb{R}$. Please refer to Salamon (2016) for more details.

$\mathbb{R}^d$ with the property that the probability of a δ sampled from $\mathbb{P}$ violates the constraint $g(x, \delta) \leq 0$ is smaller than ϵ. Unfortunately, the feasible set of (1) may be non-convex, even when the function $g(\cdot, \delta) : \mathbb{R}^d \to \mathbb{R}$ is convex for all $\delta \in \Delta$, so problem (1) is in general hard to solve. One common approach to obtain, with high probability, a feasible solution to problem (1) is by means of the scenario approach theory, which studies the optimal solution of the scenario program

$$\begin{aligned} \underset{x\in\mathcal{X}}{\text{minimise}} \quad & c^\top x \\ \text{subject to} \quad & g(x, \delta_i) \leq 0, \text{ with } \delta_i \in S \setminus R(S), \end{aligned} \tag{2}$$

where $\mathcal{X}$ and g are defined as above, with the additional assumption that $g(\cdot, \delta) : \mathbb{R}^d \to \mathbb{R}$ is convex for all $\delta \in \Delta$, $S = \{\delta_1, \ldots, \delta_m\}$ is a set of i.i.d. samples from $\mathbb{P}$, and $R(S)$ contains scenarios that have been discarded by a removal procedure. The dependence of the removed scenarios on S is made explicit.

Some concepts at the core of the scenario approach theory which will be important to our developments are presented in the sequel.

Definition 1 (Support constraints) *Consider the scenario optimization problem (2). A scenario in $S \setminus R(S)$ is said to be a support constraint (or support scenario) if its removal changes the optimal solution of (2). The set of all support constraints is called the support set of (2), and will be denoted by $\operatorname{supp}(x^\star(S))$ throughout this paper.*

Definition 2 (Fully-supported problems) *A scenario optimization problem (2) 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 $\mathbb{P}^m$, which is the measure associated to product space Δ^m.*

If scenarios are not removed (i.e., $R(S) = \emptyset$) in (2), Campi and Garatti (2008) have shown, under certain technical assumptions, an upper bound on the probability of constraint violation associated to the optimal solution of (2) that is valid for all convex optimisation problems and that holds with equality if the corresponding scenario program is fully-supported. Throughout this paper we consider the following assumption.

Assumption 1 *Problem (2) is fully-supported[3] and its solution exists and is unique for any $\{\delta_1, \ldots, \delta_m\}$. Moreover, its feasible set has a non-empty interior.*

Throughout this paper, we consider that the samples in S are ordered, i.e., there exists a bijection $\sigma : \{1, \ldots, m\} \to S$, and, for any $i, j \in \{1, \ldots, m\}$, $i \neq j$, we say that δ_i is smaller than δ_j whenever $\sigma^{-1}(\delta_i) \leq \sigma^{-1}(\delta_j)$ in the usual sense. Strict inequalities can be used with a similar interpretation. For a fixed $S = \{\delta_1, \ldots, \delta_m\}$, we also denote the optimal solution of a scenario program as in (2) for a generic $J \subset S$ as

$$\begin{aligned} z^\star(J) = \underset{x\in\mathcal{X}}{\operatorname{argmin}} \quad & c^\top x \\ \text{subject to} \quad & g(x, \delta) \leq 0, \quad \delta \in J. \end{aligned} \tag{3}$$

We now introduce the removal procedure analysed in Romao et al. (2020a). Let $r < m$ be the number of discarded constraints and write $r = q_1 d + q_2$ using the division algorithm, where q_1 and

3. Most of the results in this paper can be extended to non-fully-supported but non-degenerate problems (see Campi and Garatti (2008) for the definition) using the same technique as in Calafiore (2010) and Romao et al. (2020a), by ordering the samples and creating an augmented (regularised) optimisation problem.

P_k	Removed till $k \in \{0,\dots,q_1\}$	Optimiser at $(k+1)$-th stage
0	$R_0(S) = \emptyset$	$x_0^\star(S)$
1	$R_1(S) = \mathrm{supp}(x_0^\star(S))$	$x_1^\star(S)$
$\vdots$	$\vdots$	$\vdots$
q_1	$R_{q_1}(S) = R_{q_1-1}(S) \cup \mathrm{supp}(x_{q_1-1}^\star(S))$	$x_{q_1}^\star(S)$
q_1+1	$R_{q_1}(S) \cup \bar{R}(S)$	$x_{q_1+1}^\star(S)$

Table 1: Description of the quantities at the interim stages for the procedure encoded by (4).

q_2 are integers and $q_2 < d$. For $k \in \{0,\dots,q_1\}$, consider the sequence of q_1+1 scenario programs

$$\begin{aligned} P_k : \underset{x\in\mathcal{X}}{\text{minimise}} \quad & c^\top x \\ \text{subject to} \quad & g(x,\delta) \le 0, \quad \delta \in S \setminus R_k(S), \end{aligned} \tag{4}$$

where $R_0(S)$ is the empty set, $R_k(S) = R_{k-1}(S) \cup \mathrm{supp}(x_{k-1}^\star(S))$ for $k \in \{1,\dots,q_1\}$, with $x_k^\star(S)$, $k = 0,\dots,q_1$, representing the optimal solution of (4). If q_2 is not equal to zero, we define similarly a scenario program P_{q_1+1} with $R_{q_1+1}(S) = R_{q_1}(S) \cup \bar{R}(S)$, where $\bar{R}(S)$ is a subset of size q_2 from $\mathrm{supp}(x_{q_1}^\star(S))$ containing the q_2-th smallest scenarios according to ordering defined by σ. As the scenario program P_k depends on the solution of the previous stage through $R_k(S)$, this removal scheme can be interpreted as a cascade of q_1+2 (or q_1+1, if q_2 is equal to zero) scenarios programs where at each stage the support set associated to the optimal solution is removed and possibly a subset of the support set if $k = q_1$ and $q_2 \neq 0$. Each of these quantities are summarised in Table 1. Let

$$x^\star(S) = \begin{cases} x_{q_1}^\star(S), & \text{if } q_2 = 0; \\ x_{q_1+1}^\star(S), & \text{if } q_2 \neq 0. \end{cases} \tag{5}$$

Observe that $x^\star(S)$ is the final decision whose probability of constraint violation we are ultimately interested in.

If $q_2 = 0$, i.e., if the removed scenarios form an integer multiple of d ($r = q_1 d$), then the results in Romao et al. (2020a) allow assessing the probability of constraint violation of $x^\star(S)$. The authors in Romao et al. (2020a) also show that the bound on the probability of constraint violation is tight. To this end, Romao et al. (2020a) imposes the following assumption.

Assumption 2 *Let $S = \{\delta_1,\dots,\delta_m\}$ be i.i.d. samples from the unknown probability distribution $\mathbb{P}$ and let $C \subset S$ be any subset of S. For any $k \in \{0,\dots,q_1\}$ if $\delta \in \mathrm{supp}(x_k^\star(C))$, then we have that*

$$g(z^\star(J),\delta) > 0, \textit{ for all } J \subset C \setminus (\cup_{j=0}^{k-1}\mathrm{supp}(x_j^\star(C)) \cup \{\delta\}).$$

Remark 3 *Observe that Assumption 2 requires, roughly speaking, that scenarios removed at a given stage of the procedure are violated by the optimal solution of all scenario programs that follow it. Observe also that Assumption 2 above implies Assumption 2.2 in Campi and Garatti (2011), which requires the optimal solution of the removal procedure to violate all the removed scenarios. It is elusive at the moment how to test a priori whether a scenario program satisfies Assumption 2; however, we know that this class is non-empty by the results in Romao et al. (2020a).*

Under Assumption 2, Romao et al. (2020a) establish the following theorem.

Theorem 4 (Romao et al. (2020a)) *Consider Assumptions 1 and 2. Fix $\epsilon \in (0,1)$, set $r = q_1 d$ and let $m > r + d$. Consider also the scenario discarding scheme as encoded by (4), and note that $x^\star(S) = x^\star_{q_1}(S)$, as $q_2 = 0$. We then have that*

$$\mathbb{P}^m\Big\{(\delta_1,\dots,\delta_m) \in \Delta^m : \mathbb{P}\big\{\delta \in \Delta : g(x^\star(S),\delta) > 0\big\} > \epsilon\Big\} = \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \quad (6)$$

Note that the number of removed scenarios is restricted to be a multiple of the dimension of the optimisation problem. Besides, Romao et al. (2020a) proves that (6) holds with inequality for any non-degenerate scenario programs.

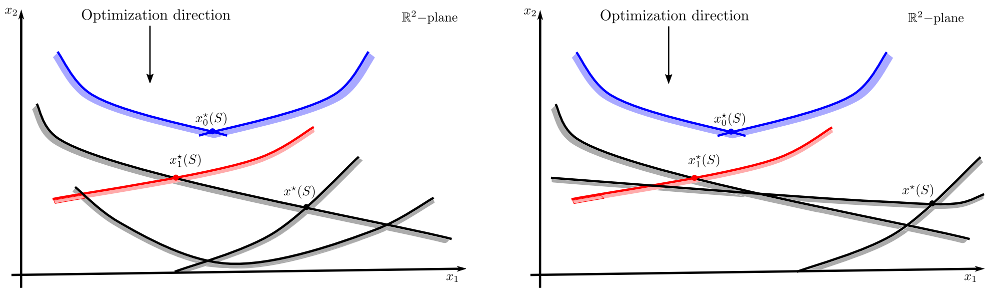
3. Main results

Throughout this section we will explore the removal strategy described by (4) when the number of discarded scenarios is not a multiple of the dimension of the optimisation problem, i.e., when $q_2 \neq 0$. We first show that this is not a straightforward generalisation of the analysis presented in Romao et al. (2020a), as it entails certain difficulties. To this end, consider two realisations of a 2-dimensional ($d = 2$) scenario program as depicted in Figure 1. In both of these realisations our goal is to remove three of the six samples, i.e., we have $q_1 = 1$ and $q_2 = 1$.

We focus first on the realisation shown in Figure 1a. Following the procedure described in (4), the 1st stage removes the scenarios highlighted in blue, as these compose the support set of $x^\star_0(S)$. To remove the third scenario we solve the corresponding scenario program without the scenarios highlighted in blue and obtain $x^\star_1(S)$ as the optimal solution. Assume that the ordering (as detailed in Section 2) is such that the scenario highlighted in red is discarded, thus leading to the solution depicted as $x^\star(S)$. For this realisation, the set composed by the two blue scenarios, the red scenario, and the support set of $x^\star(S)$ constitutes a subset of the samples with cardinality equal to $r + d = 3 + 2 = 5$ such that following the same procedure using only these 5 samples we would obtain the same solutions. Informally, this is related to the notion of compression that will be introduced in the sequel and plays a fundamental role in offering probabilistic feasibility guarantees for the returned solution. Unfortunately, this conclusion is sample dependent and does not hold uniformly across all the samples. For instance, consider the realisation illustrated in Figure 1b. The removal algorithm described in (4) proceeds similarly as in the previous case; however, we notice that the final decision is supported by two scenarios that do not belong to the support set of the previous iteration. This latter fact implies that the cardinality of the subset of the samples that would lead to the same solutions with those that would have been obtained if all the samples were employed is no longer 5 but 6. The difference between these two instances is that in the first one the support sets associated to the two last stages overlap, while in the second one these are disjoint. Moreover, the tighter the bound one can offer the smaller the cardinality of that "representative" subset of the samples. In view of a tight bound, this observation motivates restricting attention to the class of problems where the one of Figure 1a belongs. We formalise this in the next section.

3.1. Arbitrary number of removed scenarios under Assumption 2.

Inspired by the discussion in the previous section, the most natural direction if one wants to produce a tight bound on the resulting decision is that of preventing the situation of Figure 1b to happen. The main result of this section, and also of this paper, is to reveal that this can be obtained by means of the Assumption 2, which was exploited in Romao et al. (2020a) to obtain Theorem 4. To this end, the following proposition is instrumental.



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

Figure 1: Two different realisations (1a and 1b) of a two dimensional ($d = 2$) scenario program with six scenarios ($m = 6$) from which three scenarios are discarded ($r = 3$). The scenarios highlighted in blue represent the support set of the 1st stage of the removal procedure, and the ones in red the scenarios removed in the 2nd stage. In realisation 1a the scenario that has not been removed in the 2nd stage belongs to the support set of $x^\star(S)$, which is the optimal solution of the 3rd stage, while in the realisation in 1b none of the remaining scenarios from the 2nd stage belong to the support set of $x^\star(S)$.

Proposition 1 *Consider the removal procedure encoded by* (4)*. Let* $S = \{\delta_1, \ldots, \delta_m\}$ *be i.i.d. samples from the unknown probability distribution* $\mathbb{P}$*, and* $r = q_1 d + q_2$*, with* $0 < q_2 < d$*. Under Assumptions 1 and 2, if* δ *is a scenario in* $\mathrm{supp}(x^\star_{q_1}(S))$ *that has not been removed in the* (q_1+1)*-th stage, i.e.,* $\delta \in \mathrm{supp}(x^\star_{q_1}(S)) \setminus \bar{R}(S)$*, then* δ *is in the support set of* $\mathrm{supp}(x^\star(S))$*.*

Proof Consider the P_{q_1+1} that would arise if $q_2 \neq 0$ and recall that, by (5) , $x^\star(S) = x^\star_{q_1+1}(S)$. Recall also that $R_{q_1+1}(S) = R_{q_1}(S) \cup \bar{R}(S)$, where $\bar{R}(S)$ contains the q_2-th smallest scenarios of $\mathrm{supp}(x^\star_{q_1}(S))$ that will be removed at the $(q_1 + 1)$-th stage. With this in mind, let us prove this proposition by contradiction.

Suppose there exists $\bar{\delta} \in \mathrm{supp}(x^\star_{q_1}(S)) \setminus \bar{R}(S)$ that is not of support for $x^\star(S)$. Such a $\bar{\delta}$ is feasible for problem P_{q_1+1}, i.e., we must have that $g(x^\star(S), \bar{\delta}) \leq 0$. Choose $\bar{J} = \mathrm{supp}(x^\star(S)) \subset S \setminus \{R_{q_1}(S) \cup \{\bar{\delta}\}\}$ (due to the fact that $\bar{\delta} \notin \mathrm{supp}(x^\star(S))$), which then implies that $z^\star(\bar{J}) = x^\star(S)$ and $g(z^\star(\bar{J}), \bar{\delta}) \leq 0$. Under Assumption 2, with $k = q_1$ and since $R_{q_1}(S) = \cup_{j=0}^{q_1-1} \mathrm{supp}(x^\star_j(S))$, the latter is a contradiction, since this would require $g(z^\star(J), \bar{\delta}) > 0$. This concludes the proof of the proposition. ■

Proposition 1 shows that under Assumption 2 the realisation of Figure 1b can only happen with probability zero. Proposition 1 will be used as the main step to extend the results of Romao et al. (2020a). To achieve this, we follow Romao et al. (2020a) by relying on the concept of compression to establish a probably approximately correct (PAC) bound similar to that of Theorem 4.

Definition 5 (Compression set) *Let* $S = \{\delta_1, \ldots, \delta_m\}$ *be a set of i.i.d. samples from an unknown probability distribution* $\mathbb{P}$ *and* $\mathcal{A} : \Delta^m \to 2^\Delta$ *be a mapping. A subset* $C \subset S$ *with* $|C| = \zeta$ *is said to be a compression set of cardinality* ζ *associated to* $\mathcal{A}(\cdot)$ *if* $\delta \in \mathcal{A}(C)$ *for all* $\delta \in S$*.*

In other words, a compression set C contains sufficient information to generate a subset $\mathcal{A}(C)$ that contains all the samples in S. This latter property is called consistency within the learning literature (Floyd and Warmuth (1995); Vidyasagar (2002)). The notion of compression sets can be used to derive PAC bounds that quantify the confidence with which $\mathcal{A}(C)$ is an approximation of the uncertain space Δ. Not only are we interested in existence but also uniqueness of a compression set related to a mapping $\mathcal{A}(C)$, as these will allow us to characterise an exact bound on the confidence of $\mathcal{A}(C)$ as an approximation for Δ. This is summarized below.

Theorem 6 (Theorem 3, Margellos et al. (2015)) *Fix $S = \{\delta_1, \dots, \delta_m\}$ and $\epsilon \in (0, 1)$. If there exists a unique compression set of cardinality $\zeta < m$ associated to a mapping $\mathcal{A}(\cdot)$, then*

$$\mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : \delta \notin \mathcal{A}(C)\} > \epsilon\} = \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{7}$$

In the sequel, we identify, under Assumption 2, how the analysis carried in Romao et al. (2020a) can be extended to encompass an arbitrary number of discarded scenarios. We consider the removal strategy described in the previous section. Since all the intermediate problems are fully-supported we remove at each stage the associated support set and in the $(q_1 + 1)$-th stage only a subset of the support set is removed, if q_2 is not zero. Define $\mathcal{A} : \Delta^m \to 2^\Delta$ as

$$\mathcal{A}(C) = \{\delta \in \Delta : g(x^\star(C), \delta) \le 0\} \cup \left\{ \bigcup_{j=0}^{q_1-1} \mathrm{supp}(x_j^\star(C)) \cup \bigcup_{\delta \in \bar{R}(C)} \delta \right\}, \tag{8}$$

which contains the discarded scenarios in the discrete set, and the set we are ultimately interested in, namely, the set $\{\delta \in \Delta : g(x^\star(C), \delta) \le 0\}$. Following the algorithmic description presented in (4) a candidate compression set is given by

$$C = \bigcup_{j=0}^{q_1} \mathrm{supp}(x_j^\star(S)) \cup \mathrm{supp}(x^\star(S)), \tag{9}$$

as it contains, under Assumption 1, all the support sets associated to the scenario programs P_k, $k \in \{0, \dots, q_1 + 1\}$ (due to the fact that $q_2 \neq 0$ and Proposition 1 holds).

Remark 7 *By Proposition 1, we have that* $\mathrm{supp}(x_{q_1}^\star(S)) \cup \mathrm{supp}(x^\star(S)) = \bar{R}(S) \cup \mathrm{supp}(x^\star(S))$, *as any* $\delta \in \mathrm{supp}(x_{q_1}^\star(S))$ *but not in* $\bar{R}(S)$ *will be in* $\mathrm{supp}(x^\star(S))$. *As such, $|C| = r + d$ as opposed to $(q_1 + 2)d$.*

Proposition 2 *Consider the removal procedure described by* (4). *Under Assumptions 1 and 2, the set in* (9) *is the unique compression set of size $r + d$ associated to the mapping* (8).

Proof We first prove that (9) is the unique compression for (8) assuming that $x_k^\star(C) = x_k^\star(S)$, for $k \in \{0, \dots, q_1 + 1\}$.

We start by showing that C is a compression for (8). Let $\bar{\delta}$ be any scenario in S, we need to show that $\bar{\delta} \in \mathcal{A}(C)$. Note that such a $\bar{\delta}$ either belongs to the discrete part of (9), or is feasible to the problem P_{q_1+1}. In the former case, $\bar{\delta}$ is in (9) by definition. In the latter case, it is also in $\mathcal{A}(C)$ since all these scenarios are in $\{\delta : g(x^\star(S), \delta) \le 0\}$ (since $x^\star(S) = x^\star(C)$). This shows that (9) is a compression set for (8).

Before we proceed to the uniqueness proof, note that $|C| = r + d$ by Remark 7. With this in mind, let C', $C' \neq C$, be another compression of cardinality equal to $r + d$ for the mapping in (8). Let $\bar{k}$ be the minimum k for which $x_k^\star(S) = x_k^\star(C) \neq x_k^\star(C')$. Pick $\bar{\delta} \in \mathrm{supp}(x_{\bar{k}}^\star(S)) \setminus \mathrm{supp}(x_{\bar{k}}^\star(C'))$ such that $\bar{\delta} \in C \setminus C'$, such a $\bar{\delta}$ exists as otherwise we would contradict the fact that

$x_{\bar{k}}^{\star}(S) \neq x_{\bar{k}}^{\star}(C')$. A similar argument has been used in the proof of Proposition 2, item ii), in Romao et al. (2020a), inspired by Lemma 2.12 of Calafiore (2010). Hence, due to the fact that $\bar{\delta} \notin C \setminus C'$ we have that $\bar{\delta} \notin \mathrm{supp}(x_k^{\star}(C'))$, for all $k \in \{0, \dots, q_1 + 1\}$, and in particular $\bar{\delta} \notin \mathrm{supp}(x^{\star}(C'))$. Notice that $\bar{J} = \mathrm{supp}(x^{\star}(C')) \subset C' \setminus \{\bigcup_{j=0}^{\bar{k}-1} \mathrm{supp}(x_j^{\star}(S)) \cup \{\bar{\delta}\}\}$ since $x_k^{\star}(S) = x_k^{\star}(C')$ for all $k \in \{0, \dots, \bar{k} - 1\}$. By Assumption 2 this would imply that $g(z^{\star}(\bar{J}), \bar{\delta}) = g(x^{\star}(C'), \bar{\delta}) > 0$ (recall that $z^{\star}(\bar{J}) = x^{\star}(C')$), which contradicts the fact that $\bar{\delta} \in \mathcal{A}(C')$.

To conclude the proof, it remains to be shown that $x_k^{\star}(S) = x_k^{\star}(C)$ for any $k \in \{0, \dots, q_1 + 1\}$. This can be done by induction. For $k = 0$, note that $x_0^{\star}(S) = x_0^{\star}(C)$ since $\mathrm{supp}(x_0^{\star}(S)) \subset C$. Proceeding inductively, we can show the general case. We omit the details here for the sake of brevity, but we point out that this has been proved in Romao et al. (2020a), Proposition 1. We then conclude that C is the unique compression to mapping in (9). ∎

The next theorem follows trivially from Proposition 2 and Theorem 6; notice that the result below is tight, i.e., it holds with equality.

Theorem 8 *Consider the removal scheme encoded by* (4) *and suppose Assumptions 1 and 2 hold. Let $S = \{\delta_1, \dots, \delta_m\}$ be i.i.d. samples from the unknown distribution $\mathbb{P}$, $r < m$ be the number of discarded scenarios, and $\epsilon \in (0, 1)$ be given. Write $r = q_1 d + q_2$ and denote as $x^{\star}(S)$ as in* (5)*. Then we have that*

$$\mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^{\star}(S), \delta)\} > \epsilon\} = \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1 - \epsilon)^{m-i}. \qquad (10)$$

Proof The result for $q_2 = 0$ has been proved in Romao et al. (2020a). If $q_2 \neq 0$, then we have by Proposition 2 that a unique compression set exists with cardinality $r + d$. The right-hand side in (10) follows then *mutatis mutandis* from Romao et al. (2020a) with the only difference that the cardinality of the compression set is different. ∎

3.2. General removal scheme without Assumption 2

In the previous section we have extended the analysis of the removal algorithm proposed in Romao et al. (2020a) to a general number of discarded scenarios by relying on Assumption 2. In this section, we indicate possible extensions of such procedure without any further restriction on the underlying scenario program, i.e., by relaxing Assumption 2. Proofs for these statements can be found in Romao et al. (2021)

The results in Section 3.1 rely on the fact that the cardinality of the set (9) is equal to $r+d$ (which is an immediate consequence of Proposition 1). In fact, without requiring that all the remaining scenarios from the $(q_1 + 1)$-th stage are in the support set of the final solution $x^{\star}(S)$ the tight bound claimed in Theorem 4 does not hold, and the realisation depicted in Figure 1b provides one such instance. We can, however, establish a bound on the probability of constraint violation, but it will no longer be a tight one. This is described in the sequel.

Theorem 9 *Consider the removal scheme described in* (4)*. Let $S = \{\delta_1, \dots, \delta_m\}$ be i.i.d. samples from an unknown distribution $\mathbb{P}$ and r be an integer such that $m > \lceil r \rceil_d + d$, where $\lceil r \rceil_d$ is the smallest multiple of d that is larger than r. For any $\epsilon \in (0, 1)$ we have that*

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

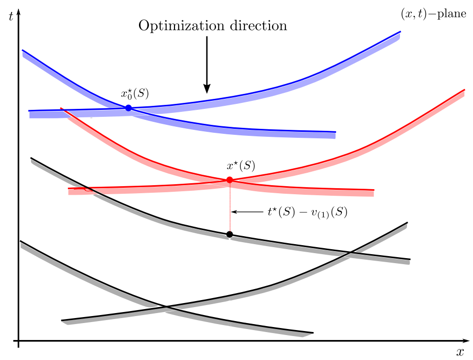


Figure 2: Alternative removal scheme suitable for min-max scenario programs with guaranteed bounds on the probability of constraint violation. We first remove scenarios by removing the support set, and then improve the cost at the last stage by moving downwards, if necessary. The blue and red scenarios correspond to first and second stages of the removal procedure. The dashed-red line defines $v_1(S)$.

The proof of Theorem 9 follows closely the ones of Theorems 3 and 4 in Romao et al. (2020a) and the proof of Theorem 7 in this paper, i.e., creating a specific mapping that involves the probability of constraint violation and showing that there exists a unique compression set of cardinality equal to $\lceil r \rceil_d + d$ associated to such a mapping.

It is worth highlighting that the bound of Theorem 9, though not tight, is strictly better than the one in Campi and Garatti (2011). Hence, the former can be satisfied, for the same probability of constraint violation and confidence level, with a larger number of removed constraints, which may then lead to better cost improvements. Besides, the computational cost of our removal strategy is often better than the cost of the greedy removal strategy, as finding the support set is cheap compared to solving a scenario program. Due to the lack of space, both of these important issues are not discussed in this paper. The reader is referred to Romao et al. (2020a) for further discussion on these aspects.

MIN-MAX SCENARIO PROGRAMS

We can now consider the class of min-max scenario programs. Let $f : \mathcal{X} \times \Delta \to \mathbb{R}$ be a function, where $\mathcal{X}$ and Δ are defined as before. Assume $f(\cdot, \delta)$ is convex for all $\delta \in \Delta$. Given m samples $S = \{\delta_1, \ldots, \delta_m\}$, we want to solve the following min-max scenario program

$$\min_{x \in \mathcal{X}} \max_{\delta \in S} f(x, \delta), \tag{11}$$

which can be cast, through an epigraphic reformulation, as the following scenario program

$$\begin{aligned} \underset{(x,t) \in \mathcal{X} \times \mathbb{R}}{\text{minimise}} \quad & t \\ \text{subject to} \quad & f(x, \delta) \leq t, \text{ for all } \delta \in S. \end{aligned} \tag{12}$$

Consider the following removal scheme, which is inspired by the results in Care et al. (2015) and Garatti et al. (2019) where the empirical cost for min-max scenario programs is characterised. For a given positive integer r such that $m > r + d + 1$ we proceed similarly as in the procedure described by (4), i.e., writing $r = q_1(d + 1) + q_2$ and removing $(q_1 + 1)d$ scenarios by means of a cascade of scenarios programs in which the support set is removed at each stage. However, at the (q_1+1)-th stage, rather than choosing a subset of size q_2 from $\mathrm{supp}((x^\star_{q_1}(S), t^\star_{q_1}(S)))$ to be discarded we compute the quantity

$$v_i(S) = t^\star_{q_1}(S) - f(x^\star_{q_1}(S), \delta_i), \text{ for all } \delta_i \in S \setminus \{R_{q_1}(S) \cup \mathrm{supp}((x^\star_{q_1}(S), t^\star_{q_1}(S)))\},$$

where $(x^\star_k(S), t^\star_k(S))$, $k \in \{0, \dots, q_1\}$, is the optimal solution of the scenario program (12), treated as a particular instance of the scenario program (4). It is important to notice that each $v_i(S)$ corresponds to the vertical distance between $t^\star_{q_1}(S)$ and the intersection of the constraint generated by the i-th scenario with the vertical line that passes through $x^\star_{q_1}(S)$ (see Figure 2 for an illustration). We then pick the q_2-th smallest $v_i(S)$ and denote them as $v_{(1)}(S) < v_{(2)}(S) < \ldots < v_{(q_2)}(S)$. The q_2-th layer probability of constraint violation associated to the optimal solution of (12) is then given by (denoting $x^\star_{q_1}(S) = x^\star(S)$ and $t^\star(S) = t^\star_{q_1}(S)$)

$$V_{q_2}(S) = \mathbb{P}\{\delta \in \Delta : f(x^\star(S), \delta) > t^\star(S) - v_{(q_2)}(S)\}, \tag{13}$$

which constitutes the probability that an unseen sample has a cost greater than $t^\star(S) - v_{(q_2)}(S)$. An illustration of this procedure for $d = 1, r = 3$, and $m = 9$ is depicted in Figure 2. Under this setting, we can state the following theorem.

Theorem 10 *Consider the removal scheme described in this section and let $V_{q_2}(S)$ be defined as in (13). Let $S = \{\delta_1, \dots, \delta_m\}$ be i.i.d. samples from an unknown distribution $\mathbb{P}$ and r be an integer such that $m > r + d + 1$. If the min-max scenario program (11) admits a unique solution, the for any $\epsilon \in (0, 1)$ we have that*

$$\mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : V_{q_2}(S) > \epsilon\} \le \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}.$$

4. Conclusion

We have analysed how a removal procedure within the sampling-and-discarding approach that yields tight results on the probability of constraint violation of the resulting solution can be extended to an arbitrary number of discarded constraints. Previous results in the literature restricted the number of discarded scenarios to be a multiple of the dimension of the optimisation problem.

In this paper, we have shown that under Assumption 2 the constraint on the number of removed scenarios can be lifted. However, it is elusive how large the class of scenario programs satisfying such an Assumption 2 is. To overcome such a shortcoming we discuss two bounds that allow the decision maker to trade feasibility and performance without requiring further assumptions on the scenario programs. The first one considers the fact that support scenarios may not be shared in the final stage of the procedure, thus leading to a conservative estimate on the resulting probability of violation, which, however, holds uniformly for all samples in Δ^m and for an arbitrary number of discarded scenarios. The second bound holds for the so-called min-max scenario programs and combines our removal strategy with the results in Care et al. (2015) and Garatti et al. (2019).

Acknowledgments

This work was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES), Ministry of Education, Brazil, and by the EPSRC UK under grants EP/P03277X/1 and EP/M002454/1.

References

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

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

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

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

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.

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

Peyman Esfahani, Tobias Sutter, and John Lygeros. Performance Bounds for the Scenario Approach and an Extension to a Class of Non-Convex Programs. *IEEE Transactions on Automatic Control*, 60(1):46–58, 2015.

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

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

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.

Arkadi Nemirovski and Alexander Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):959–996, 2006.

Bernardo Pagnoncelli, Shabbir Ahmed, and Alexander Shapiro. Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009.

András Prékopa. Probabilistic programming. In *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267 – 351. Elsevier, 2003.

Licio Romao, Kostas Margellos, and Antonis Papachristodoulou. On the exact feasibility of convex scenario programs with discarded constraints. *arXiv:2003.13433 [math.OC]*, pages 1–15, 2020a.

Licio Romao, Kostas Margellos, and Antonis Papachristodoulou. Tight generalization guarantees for the sampling and discarding approach to scenario optimization. In *59th Conference on Decision and Control. To appear*, 2020b.

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. URL http://sysos.eng.ox.ac.uk/wiki/images/4/46/RMP20eL4DC.pdf.

Dietmar A. Salamon. *Measure and Integration*. European Mathematical Society, 2016, Zurich, Switzerland, 2016.

M. Vidyasagar. *Learning and Generalization*. Springer, 2002.