# On the exact feasibility of convex scenario programs with discarded constraints

Licio Romao, Antonis Papachristodoulou, and Kostas Margellos

*Abstract*—We revisit the so-called sampling and discarding approach used to quantify the probability of constraint violation of a solution to convex scenario programs when some of the original samples are allowed to be discarded. Motivated by two scenario programs that possess analytic solutions and the fact that the existing bound for scenario programs with discarded constraints is not tight, we analyze a removal scheme that consists of a cascade of optimization problems, where at each step we remove a superset of the active constraints. By relying on results from compression learning theory, we show that such a removal scheme leads to less conservative bounds for the probability of constraint violation than the existing ones. We also show that the proposed bound is tight by characterizing a class of optimization problems that achieves the given upper bound. The performance improvement of the proposed methodology is illustrated by an example that involves a resource sharing linear program.

*Index Terms*—Scenario approach, randomized algorithms, chance-constrained optimization, probabilistic methods.

## I. INTRODUCTION

Uncertain optimization programs capture a wide class of engineering applications. Tractability of this class of optimization problems is an active area of research [1]–[8]. In the last decades, several approaches have been developed to cope with uncertainty in an optimization context. Among those, robust optimization [9]–[12] has been successfully applied to several control problems [13]–[18]. It consists of making certain assumptions, often arbitrary, on the geometry of the uncertainty set (ellipsoidal, polytopic, etc.) and then optimizing over the worst case performance within this set. Another approach is chance-constrained optimization [19]–[21] that relies on imposing constraints that only need to be satisfied with given probability. However, these problems are hard to solve in general, without imposing any assumption on the underlying distribution of the uncertainty (e.g., Gaussian).

An alternative to robust and chance-constrained optimization involves data driven algorithms. Within this context, this paper lies in the realm of the scenario approach theory [8], [22]–[29]: a randomized technique which involves generating a finite number of scenarios and enforcing a different constraint for each of them. Under convexity, the optimal solution to such a scenario program is shown to be feasible (with certain probability) to the associated chance-constrained program. One of the fundamental developments in the scenario approach literature is to provide a distribution-free result that holds for all convex problems [24]. Moreover, the bound proposed in [24] is tight in the sense that it is achieved by the so-called class of fully-supported optimization problems, however, it might be conservative for more general problem classes.

To alleviate such conservatism and trade feasibility to performance, the so-called sampling and discarding [25] (see also [28]) was introduced; a similar result known as scenario approach with constraint removal was also developed in [26]. These allow removing some of the extracted scenarios and enforcing the constraints only on the remaining ones, thus improving the performance in terms of optimality of the resulting solution. As opposed to the original bound in [24], however, the bound on the probability of constraint violation in [25], [26] is not tight.

Similarly to the motivation of [25], our main goal is to improve performance and decrease the conservatism of the solution obtained by means of the scenario approach theory. We capitalize on the fact that the bound of the sampling-and-discarding scheme is not tight to provide a less conservative and tight bound on the probability of constraint violation for convex scenario programs with discarded constraints. To this end, we develop a novel analysis to study a removal procedure that constitutes solving a cascade of scenario programs and removing, at each stage, scenarios in an integer multiple of the dimension of the decision variables. Our theoretical findings bear important consequences in the application of the scenario theory to control problems [23], [30]–[40], as we may be able to achieve better performance while guaranteeing the same level of constraint violation and confidence. The proposed bound on the probability of constraint violation is similar in terms of complexity to the one of [25], [26]; it is also distribution-free and holds, under a non-degeneracy assumption (to be formally defined in the sequel), for all convex problems. We also show that the resulting bound is tight, and characterize the class of scenario programs for which this is the case. As such, our results extend the ones of [25], [26] and cannot be further improved. To summarize, our main contributions are:

- Proposing and analyzing a removal scheme that possesses tighter guarantees than [25], [26] on the probability of constraint violation for scenario programs with discarded constraints (Theorems 3 and 4).
- Proving tightness of the resulting bound by characterizing

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

the class of scenario programs that satisfies our bound with equality (Theorem 5).

- Computing analytically the solution of two scenario programs where the bound is tight (Section III).
- Relaxing an assumption present in [25] that requires the removed scenarios to be violated by the final solution.
- Developing a novel proof line. Our analysis departs from the one of [25], and is based on probably approximately correct (PAC) learning concepts that use the notion of compression [27], [41], [42].

It is important to highlight that our analysis holds for a particular discarding scheme, which requires removing scenarios in batches, preventing us to remove them one by one. Extension to this direction is outside the scope of the current paper. Moreover, all our results are *a priori*; possibly less conservative but *a posteriori* results are available [7], [8], [29], however, follow a different conceptual and analysis line from the one adopted in this paper.

The paper is organized as follows: Section II reviews some background results on the scenario approach with discarded constraints and certain learning theoretic concepts. Section III motivates the main results of the paper by means of two scenario programs that possess analytic solutions. Section IV introduces the proposed scenario discarding scheme and states the main results of the paper. Proofs are provided in Section V. Section VI characterizes the class of optimization programs for which the proposed result is tight and Section VII illustrates the theoretical results by means of a numerical example. Finally, Section VIII concludes the paper and provides some directions for future work.

## II. SCENARIO OPTIMIZATION WITH DISCARDED SCENARIOS

We first review existing results on the so-called sampling and discarding approach for scenario based optimization [25], [26]. We then present some preliminary learning theoretic concepts that are crucial for our developments.

### A. Sampling and discarding

Let $\Delta$ be the space where an uncertainty vector takes values from and denote by $(\Delta, \mathcal{F}, \mathbb{P})$ the associated probability space, where $\mathcal{F}$ is a $\sigma$-algebra and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure on $\Delta$ (see [43] for more details). Fix any $m \in \mathbb{N}$, and let $S = \{\delta_1, \delta_2, \ldots, \delta_m\}$ be independent and identically distributed (i.i.d.) samples from $\mathbb{P}$. Note that $(\delta_1, \ldots, \delta_m) \in \Delta^m$; a natural probability space associated with $\Delta^m$ is $(\Delta^m, \otimes_{i=1}^m \mathcal{F}, \mathbb{P}^m)$, where $\otimes_{i=1}^m \mathcal{F}$ is the smallest $\sigma$-algebra containing the cone sets $\prod_{i=1}^m \mathcal{F}$. Our analysis is based on a data-driven interpretation, where $\mathbb{P}$ is considered to be fixed, but possibly unknown, and the only information about uncertainty is a collection of i.i.d. scenarios $S$.

We consider convex optimization programs affected by uncertainty $\delta$, and represent uncertainty by means of scenarios. This gives rise to the so-called convex scenario programs, where constraints are enforced only on the scenarios in $S$ [22],

[24], [26]. We are particularly interested in the case where some of the scenarios are removed, in view of improving the performance of the obtained solution. This is known as *sampling and discarding* in the terminology of [25] (also known as scenario approach with constraint removal in [26]).

To this end, for any set $R \subset S$, with $|R| = r < m$, consider the following problem

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

where $x \in \mathbb{R}^d$, $X$ is a closed and convex set of $\mathbb{R}^d$, and function $g : \mathbb{R}^d \times \Delta \to \mathbb{R}$ is convex in $x$ for all $\delta \in \Delta$. The subset $R$ contains scenarios that have been removed by means of a procedure that uses $S$ as input; hence, strictly speaking $R$ depends on the scenarios $S$ but this dependency is omitted for simplicity. If $R = \emptyset$, then one recovers the standard scenario approach [23], [24]. Moreover, the objective function is taken to be affine without loss of generality; in case of an arbitrary convex objective function, an epigraphic reformulation would render the problem in the form of (1). Note that only convex scenarios programs will be considered, as in [25], [26]. We impose the following assumption.

**Assumption 1** (Feasibility, Uniqueness). *For any* $S \subset \Delta^m, R \subset S$, *the optimal solution of* (1) *exists and is unique.*

In case of multiple solutions a convex tie-break rule could be selected to single-out a particular one, thus relaxing the uniqueness requirement of Assumption 1.

Denote by $x^\star(S)$ the (unique under Assumption 1) minimizer of (1). Note that we introduce $S$ as argument since the optimal solution of (1) is a random variable that depends on all extracted scenarios, i.e., it is a random variable that takes values on the space $\Delta^m$. The following result from [25] characterizes the probability that $x^\star(S)$ violates the constraints for a new realization of $\delta$ exceeds a given level $\epsilon \in (0, 1)$.

**Theorem 1** (Theorem 2.1, [25], or Theorem 4.1, [26]). *Consider Assumption 1, and fix* $\epsilon \in (0, 1)$. *Let* $m > d + r$ *and denote by* $x^\star(S)$ *the optimal solution of* (1). *If all removed scenarios are violated by the resulting solution* $x^\star(S)$*, i.e.,* $g(x^\star(S), \delta) > 0$ *for all* $\delta \in R$*, with* $\mathbb{P}^m$-*probability one, then*

$$\mathbb{P}^m \Big\{ (\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\big\{ \delta \in \Delta : g(x^\star(S), \delta) > 0 \big\} > \epsilon \Big\}$$

$$\leq \binom{r+d-1}{r} \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1 - \epsilon)^{m-i}. \tag{2}$$

Theorem 1 represents an important generalization of the scenario approach theory, as it allows the decision maker to trade feasibility to performance. Indeed, observe that the feasible set of (1) is enlarged when $R$ is non-empty (i.e., when scenarios are discarded), thus leading to a cost improvement with respect to the case where $R$ is the empty set. This fact and the bound of Theorem 1 enable the decision maker to improve cost, while controlling the probability of constraint violation.

It should be also noted that Theorem 1 does not allow for an arbitrary discarding scheme; it rather requires that, with $\mathbb{P}^m$-probability one, all discarded scenarios are violated by the resulting solution $x^\star(S)$. This is instrumental in the proof of Theorem 1, as shown in [25].

Besides, if $r = 0$ the bound in Theorem 1 is known to hold with equality for a class of scenario programs called fully-supported programs (see Section II or [24], [25] for more details). When scenarios are discarded, i.e., when $r \neq 0$, it is elusive how to construct a removable scheme that allows for a tight bound. It is shown in Section 4.2 of [25] that there exists a class of convex optimization programs and a discarding scheme such that the right-hand side of (2) can be replaced by $\sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$, however, the argument is not constructive and is limited to an existential statement, as it does not specify the underlying class nor the discarding mechanism. Moreover, this tighter bound is achieved for some probability measure which is not specified, and not in a distribution free manner.

In this paper we provide a tight bound to scenario programs when $r \neq 0$. In particular, we provide a constructive discarding mechanism that is distribution free and admits the aforementioned bound of the existential statement in [25]. To achieve this, in Section IV we introduce an alternative discarding strategy that possesses tighter guarantees for its final solution than the one of Theorem 1, at the expense of requiring the number of removed scenarios to be an integer multiple of the number of decision variables. As a byproduct of our analysis, we also relax the assumption of Theorem 1 that requires all the removed scenarios to be violated by the final solution.

### B. Learning theoretic concepts

The following definition is crucial for the results in this paper.

**Definition 1** (Compression set). *Fix $m \in \mathbb{N}$, and consider $S \subset \Delta$ with $|S| = m$. Let $\zeta < m$, and $C \subset S$ with cardinality $|C| = \zeta$. Consider a mapping $\mathcal{A} : \Delta^m \to 2^\Delta$. If*

$$\delta \in \mathcal{A}(C), \quad \text{for all } \delta \in S,$$

*then $C$ is called a compression set of cardinality $\zeta$ for $\mathcal{A}$.*

In other words, a compression set $C$ is a subset of the samples $S$ such that $\mathcal{A}(C)$, i.e., the set generated using only $\zeta$ of the samples, contains all samples in $S$, even the ones that were not included in $C$. In statistical learning theory this property is known as consistency of $\mathcal{A}(C)$ with respect to the samples [27], [41]. The main focus within a probably approximately correct (PAC) learning framework (see [41] and references therein) is to quantify the probability that $\mathcal{A}(C)$ differs from $\Delta$. Since $\mathcal{A}(C)$ depends on the scenarios in $S$ (as $C$ is a selection among all scenarios), this probability is itself a random variable defined on the product probability space $\Delta^m$.

To address this question we will use tools from PAC learnability related to compression learning. To this end, we adapt the main concepts and result of [27] to the notation of our paper.

**Theorem 2** (Theorem 3, [27]). *Fix $\epsilon \in (0,1)$ and $\zeta < m$. If there exists a unique compression set $C$ of cardinality $\zeta$, then*

$$\mathbb{P}^m \left\{ (\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : \delta \notin \mathcal{A}(C)\} > \epsilon \right\}$$

$$= \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \quad (3)$$

For a fixed $\epsilon \in (0,1)$, observe that the right-hand side of (3) goes to zero as $m$ tends to infinity. This is a desirable property, as it indicates that $\Delta$ can be asymptotically approximated by $\mathcal{A}(C)$. Moreover, for a fixed $m \in \mathbb{N}$, the result of Theorem 2 provides a non-asymptotic result, quantifying the measure of the set $\Delta \setminus \mathcal{A}(C)$. A mapping with these properties is called PAC within the learning literature. Theorem 2 states that if a mapping possesses a unique compression set, then it is at least $(1 - \epsilon)$-accurate as an approximation of $\Delta$ (approximately correct), with confidence (probably) equal to $1 - \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$.

### III. MOTIVATING EXAMPLE: TWO SCENARIO PROGRAMS WITH ANALYTIC SOLUTIONS

To motivate the bound that will be proposed in the sequel, we first provide two scenario programs that admit analytic solutions.

#### A. One-dimensional example

Suppose that $m$ i.i.d. samples, $S = \{\delta_1, \ldots, \delta_m\}$, are drawn from a uniform distribution on the interval $[0,1]$. Consider the following scenario program that is in the form of (1).

$$\begin{aligned} &\underset{x \in [0,1]}{\text{minimize}} && x \\ &\text{subject to} && x \geq \delta, \quad \delta \in S \setminus R. \end{aligned} \quad (4)$$

Under the choice of a uniform distribution, the optimal solution of (4) is unique with $\mathbb{P}^m$-probability one. Let $r < m$ be the number of discarded scenarios and consider a (natural) removal scheme that discards scenarios one by one by means of a cascade of scenario programs where at each stage the scenario corresponding to the active constraint is removed from the set $S$. For instance, the first discarded scenario, which can be explicitly computed as $\delta^{(1)} = \text{argmax}_{\delta \in S} \delta$, corresponds to the active constraint of (4) when all scenarios in $S$ are enforced. We then solve (4) with all but the scenario removed in the previous stage being enforced, thus resulting in the scenario $\delta^{(2)} = \text{argmax}_{\delta \in S \setminus \delta^{(1)}} \delta$ to be discarded. We proceed similarly until $r$ scenarios are removed.

Let $x_k^\star(S)$ be the optimal solution at the $(k+1)$-th stage of the removal procedure described in the previous paragraph. Note that $x_k^\star(S) = \delta^{(k+1)}$, $k = 0, \ldots, r$, where $\delta^{(k+1)}$ represents the $(k+1)$-th largest sample of $S$. Our goal is to compute the probability of constraint violation associated to the optimal solution of (4) when the scenarios that belong to $S \setminus \{\delta^{(1)}, \ldots, \delta^{(r)}\}$ are enforced. Since $\mathbb{P}$ is a uniform probability measure on $[0,1]$, it is clear that, for each $k \in \{0, \ldots, r\}$, such a probability is given by

$$V(x_k^\star(S)) = \mathbb{P}\{\delta \in \Delta : x_k^\star(S) < \delta\} = 1 - x_k^\star(S). \quad (5)$$

Observe that

$$\mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : V(x_k^\star(S)) > \epsilon\}$$
$$= \mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : \delta^{(k+1)} < 1 - \epsilon\}$$
$$= \sum_{i=0}^{k} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{6}$$

The first equality in (6) follows from the fact that $x_k^\star(S) = \delta^{(k+1)}$ and the second by partitioning the space $\Delta^m$ into $k+1$ disjoint sets where each of these sets contains elements $S$ for which exactly $i$, $i = 0, \ldots, k$, samples lie within the interval $[1 - \epsilon, 1]$ and then applying the total law of probability. An alternative explanation using order statistics can be found in [44]. Hence, the distribution of the probability of constraint violation associated with the final solution of the considered removal strategy is given by $V(x_r^\star(S))$, which is obtained from (6) by substituting $k = r$.

### B. Two-dimensional example

Suppose that $m$ i.i.d. samples are drawn from a uniform distribution on the interval $[0, 1]$ and consider the scenario program that returns the minimum width interval containing the samples given by

$$\begin{aligned} \underset{x,y \in [0,1],\ y \geq x}{\text{minimize}} \quad & y - x \\ \text{subject to} \quad & \delta \in [x, y], \quad \delta \in S \setminus R. \end{aligned} \tag{7}$$

For any collection of the samples $S = \{\delta_1, \ldots, \delta_m\}$, with $\mathbb{P}^m$-probability one, the scenario program (7) has a unique solution and, at the optimal solution, there are exactly two active constraints, namely, those associated with the smallest and the largest sample of $S \setminus R$. Let $r = 2\ell < m$, for some integer $\ell$, be an even integer and consider (similar as in the one dimensional example) a removal scheme that discards the active constraints of (7) at each stage. Denote by $x_k^\star(S)$, $k = 0, \ldots, \ell$, the two dimensional vector containing the optimal solution of the $(k+1)$-th stage. One of the components of $x_k^\star(S)$ is the $(k+1)$-th largest sample of $S$, which we denote by $\delta^{(k+1)}$, and the other the $(k+1)$-th smallest sample of $S$, which we denote by $\delta_{(m-k)}$.

We are interested in the probability of constraint violation at the $(k+1)$-th stage, which is given by

$$V(x_k^\star(S)) = \mathbb{P}\{\delta \in \Delta : \delta \notin [\delta_{(m-k)}, \delta^{(k+1)}]\}$$
$$= \mathbb{P}\{\delta \in \Delta : \delta < \delta_{(m-k)}\} + \mathbb{P}\{\delta \in \Delta : \delta > \delta^{(k+1)}\}$$
$$= 1 - (\delta^{(k+1)} - \delta_{(m-k)}) = 1 - L_{(k+1)}(S), \tag{8}$$

where $L_{(k+1)}(S) = (\delta^{(k+1)} - \delta_{(m-k)})$ represents the length of the interval after the removal of $2k$ samples. Equation (8) consists in the probability that a new sample is drawn from $\mathbb{P}$ and it falls outside the interval $[\delta_{(m-k)}, \delta^{(k+1)}]$. Let $\epsilon \in [0, 1]$, we have that

$$\mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : V(x_k^\star(S)) > \epsilon\}$$
$$= \mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : L_{(k+1)}(S) < 1 - \epsilon\}. \tag{9}$$
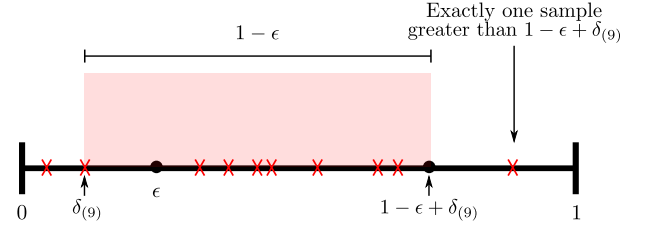


Fig. 1: Realization of a sample $S = \{\delta_1, \ldots, \delta_{10}\}$ that belongs to the set $A_1 \cap B_1 \cap E_1^1$, where its components are defined in (10) and (13).

For each $k \in \{0, \ldots, \ell\}$, let

$$A_k = \{(\delta_1,\ldots,\delta_m) \in \Delta^m : \delta_{(m-k)} \leq \epsilon\}$$
$$B_k = \{(\delta_1,\ldots,\delta_m) \in \Delta^m : V(x_k^\star(S)) > \epsilon\}, \tag{10}$$

where $A_k$ contains the samples $S$ whose $(k + 1)$-th smallest element lies in the interval $[0, \epsilon]$ and $B_k$ contains the samples that lead to $V(x_k^\star(S)) > \epsilon$. Using this notation, we can write (9) as

$$\mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : V(x_k^\star(S)) > \epsilon\} = \mathbb{P}^m\{B_k\}$$
$$= \mathbb{P}^m\{A_k^c \cap B_k\} + \mathbb{P}^m\{A_k \cap B_k\} \tag{11}$$

where $A^c$ stands for the set complement of $A$. Let us analyze each of the terms in the right-hand side of (11) separately. We start with the first term. Note that

$$\mathbb{P}^m\{A_k^c \cap B_k\} = \mathbb{P}^m\{A_k^c\}\mathbb{P}^m\{B_k|A_k^c\} = \mathbb{P}^m\{A_k^c\}$$
$$= \mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : \delta^{(k+1)} \leq 1 - \epsilon\}$$
$$= \sum_{i=0}^{k} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}, \tag{12}$$

where the second equality follows from the fact that $\mathbb{P}\{B_k|A_k^c\}$ is equal to one due to $A_k^c \subset B_k$, i.e., the length of $[\delta_{(m-k)}, \delta^{(k+1)}]$ is less than $1 - \epsilon$ whenever the $(k+1)$-th smallest sample in $S$ is larger than $\epsilon$; and the third equality due to the fact that

$$\mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : \delta_{(m-k)} > \epsilon\}$$
$$= \mathbb{P}^m\{(\delta_1,\ldots,\delta_m) \in \Delta^m : \delta^{(k+1)} \leq 1 - \epsilon\},$$

which can be obtained by simple algebraic manipulations. Finally, the last equality holds due to (6).

To compute the second term in the right-hand side of (11), it is convenient to define the partition of $\Delta^m$ as

$$E_k^j = \{(\delta_1,\ldots,\delta_m) \in \Delta^m : \text{ there are exactly } j \text{ samples}$$
$$\text{greater than } 1 - \epsilon + \delta_{(m-k)}\}, \tag{13}$$

where $j = 0, \ldots, m - k$. Note that there can be no more than $m - k$ samples greater than $\delta_{(m-k)}$, as this would contradict the fact that $\delta_{(m-k)}$ is the $(k + 1)$-th smallest sample of $S$; hence, we have that $\Delta^m = \cup_{j=0}^{m-k} E_k^j$. For instance, Figure 1

depicts a realization that belongs to $E_1^1$ when $m = 10$. The partition given in (13) allows us to write

$$\mathbb{P}^m\{A_k \cap B_k\} = \sum_{j=0}^{m-k} \mathbb{P}^m\{A_k \cap B_k \cap E_k^j\}$$

$$= \sum_{j=0}^{k} \mathbb{P}^m\{A_k \cap B_k \cap E_k^j\} + \sum_{j=k+1}^{m-k} \mathbb{P}^m\{A_k \cap B_k \cap E_k^j\}$$

$$= \sum_{j=0}^{k} \mathbb{P}^m\{A_k \cap B_k \cap E_k^j\}, \tag{14}$$

where the last equality follows from the fact that $A_k \cap B_k \cap E_k^j = \emptyset$ for all $j > k$, since having more than $k$ samples greater than $1 - \epsilon + \delta_{(m-k)}$ with $\delta_{(m-k)} \leq \epsilon$ implies that $L_{(k+1)} \geq 1 - \epsilon$, so such a realization does not belong to $B_k$.

We claim that

$$\mathbb{P}^m\{A_k \cap B_k \cap E_k^j\}$$
$$= \binom{m}{j+k+1} \epsilon^{j+k+1}(1-\epsilon)^{m-j-k-1}. \tag{15}$$

Consider first the case where $k = j = 1$. We can compute $\mathbb{P}\{A_1 \cap B_1 \cap E_1^1\}$ as

$$\mathbb{P}^m\{A_1 \cap B_1 \cap E_1^1\} = \int_0^\epsilon \mathbb{P}^m\{B_1 \cap E_1^1 | \delta_{(m-1)} = \xi\} d\xi$$

$$= \int_0^\epsilon \underbrace{(m-1)(m-2)\xi(\epsilon - \xi)}_{T_1} \underbrace{(1-\epsilon)^{m-3}}_{T_2} d\xi,$$

where the term $T_1$ is due to the fact that there must be exactly one sample less than $\xi$ and exactly one sample greater than $1 - \epsilon + \xi$ and there are $(m-1)(m-2)$ possible such samples, and the term $T_2$ is due to the fact that the remaining samples $(m-3$ in this case) is within the interval $[\xi, 1 - \xi + \epsilon]$. Please refer to Figure 1 for a realization of a sample $S$ that belongs to $A_1 \cap B_1 \cap E_1^1$ when $m = 10$.

To show (15) in general one may proceed inductively; alternatively, we can use the fact that the uniform distribution assigns the same probability to subsets of $[0, 1]$ that have the same length. Hence, $\mathbb{P}^m\{A_k \cap B_k \cap E_k^j\}$ is the probability that $j + k + 1$ samples are outside the interval $[\delta_{(m-k)}, 1 - \epsilon + \delta_{(m-k)}]$, and the remaining $m - j - k - 1$ ones to its complement. This immediately yields (15).

Combining (11), (12), (14) and (15), we have that

$$\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : V(x_k^\star(S)) > \epsilon\} = \mathbb{P}^m\{B\}$$

$$= \sum_{i=0}^{k} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$$

$$+ \sum_{j=0}^{k} \binom{m}{j+k+1} \epsilon^{j+k+1}(1-\epsilon)^{m-j-k-1}$$

$$= \sum_{i=0}^{k} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i} + \sum_{i=k+1}^{2k+1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$$

$$= \sum_{i=0}^{2k+1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{16}$$

Set $k = \ell$ and recall from the discussion following equation (7) that $r = 2\ell$, so we have proved that

$$\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : V(x_\ell^\star(S)) > \epsilon\}$$
$$= \sum_{i=0}^{r+1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \tag{17}$$

These two examples show that the associated probability of constraint violation for the case where scenarios are discarded can be computed analytically and that the obtained probability is better than the one of Theorem 1. Motivated by this fact we will show a tighter bound on the probability of constraint violation for such scenario programs that is valid for an arbitrary $d$, as long as the number of discarded scenarios is an integer multiple of $d$. The only difference in a $d$-dimensional example would be that the upper limit in the summation would be $r + d - 1$ (See Theorem 5), which is consistent with (6) and (17) and is tighter than the bound of Theorem 1. This is due to the fact that these scenario programs satisfy Assumption 4 (to be defined in the sequel), which is a sufficient condition to obtain such a tight bound.

## IV. PROPOSED DISCARDING SCHEME AND MAIN RESULTS

In this section, we formalize a removal scheme that results in a better bound on the probability of constraint violation of a scenario program with discarded constraints. For a given set of scenarios $S = \{\delta_1, \ldots, \delta_m\}$, we solve a cascade of $\ell + 1$ optimization programs denoted by $P_k$, $k \in \{0, \ldots, \ell\}$, where $(\ell + 1)d < m$. For each $k \in \{1, \ldots, \ell\}$, let

$$P_k : \quad \underset{x \in X \subset \mathbb{R}^d}{\text{minimize}} \quad c^\top x$$

$$\text{subject to} \quad g(x, \delta) \leq 0, \quad \text{for all } \delta \in S \setminus \bigcup_{j=0}^{k-1} R_j(S),$$

where $R_k(S)$, with $|R_k(S)| = d$, represents the set of removed scenarios at stage $k$, and $\bigcup_{j=0}^{k-1} R_j(S)$ the ones that have been removed up to stage $k$. For $k = 0$, we solve problem $P_0$ by enforcing all the scenarios in $S$. Notice that the number of removed scenarios at stage $\ell$ is given by $\ell d$ (the samples in the set $\bigcup_{j=0}^{\ell-1} R_j(S)$) and that, by construction, the collection of removed scenarios is disjoint. The choice of each set of discarded scenarios depends on the initial set $S$, thus we introduce it as an argument in $R_k$. A schematic illustration of the proposed scheme is provided in Figure 2. Our choice for $R_k(S)$, $k \in \{0, \ldots, \ell-1\}$, will be detailed in the following two subsections.

**Definition 2** (Support set; see Definition 2 in [24]). *Fix any $k \in \{0, \ldots, \ell\}$ and consider $P_k$. An element of $S \setminus \bigcup_{j=0}^{k-1} R_j(S)$ is a support scenario of $P_k$, if its removal changes the minimizer $x_k^\star(S)$. The support set of $x_k^\star(S)$, denoted by $\mathrm{supp}(x_k^\star(S))$, is the collection of support scenarios of $S \setminus \bigcup_{j=0}^{k-1} R_j(S)$.*
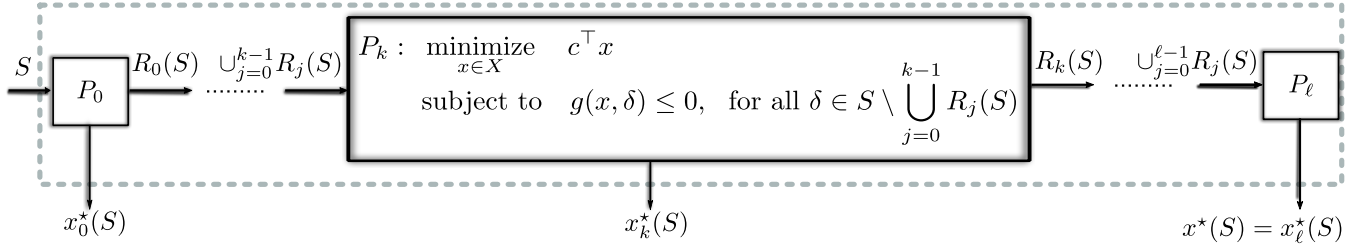
Fig. 2: Block diagram of the proposed scheme. Given $S = \{\delta_1, \ldots, \delta_m\}$, with $(\ell+1)d < m$, we solve a cascade of $\ell+1$ optimization programs denoted by $P_k$, $k \in \{0, \ldots, \ell\}$, and remove $R_k(S)$ scenarios with $|R_k(S)| = d$ at each stage. In total $r = \ell d$ scenarios (the ones in $\bigcup_{j=0}^{\ell-1} R_j(S)$) are discarded. The set of discarded scenarios depends on the initial set $S$, thus we introduce it as argument of $R_k$. If each problem is fully-supported, then $R_k(S)$ corresponds to the (unique) support set associated with the minimizer $x_k^\star(S)$ of that program – see (18); otherwise, $R_k(S)$ contains the support scenarios as well as additional scenarios selected according to a lexicographic order, as in (21). The final solution is denoted by $x^\star(S) = x_\ell^\star(S)$.

**Definition 3** (Fully-supported programs; see Definition 3 in [24])**.** *Fix any $k \in \{0, \ldots, \ell\}$ and consider $P_k$. We say that $P_k$ is fully-supported if, for any $S$ with $|S| = m$ and $m > d$, $|\text{supp}(x_k^\star(S))| = d$ with $\mathbb{P}^m$-probability one.*

**Definition 4** (Non-degenerate programs; see Assumption 2 in [8])**.** *Fix any $k \in \{0, \ldots, \ell\}$ and consider $P_k$. We say that $P_k$ is non-degenerate if, with $\mathbb{P}^m$-probability one, solving the problem by enforcing the constraints only on the support set, $\text{supp}(x_k^\star(S))$, results in $x_k^\star(S)$, i.e., the solution obtained when all samples in $S \setminus \bigcup_{j=0}^{k-1} R_j(S)$ are employed.*

Note that if a problem is fully-supported then it is also non-degenerate, however, the opposite implication does not hold. Moreover, in a convex optimization context, non-degeneracy is a relatively mild assumption, and implies that scenarios give rise to constraints at general positions that do not have accumulation points. On the contrary, requiring a problem to be fully-supported is much stronger, however, it exhibits interesting theoretical properties as the number of support scenarios is exactly equal to $d$ [24], [27].

In the sequel, we split our analysis into fully-supported and non-degenerate scenario programs. This facilitates our analysis as our proof construction is laid out better if we analyze the proposed removal scheme assuming that the scenario programs are fully-supported. In Section IV-B, we lift this assumption and show how to extend the developed analysis to the more general case of non-degenerate scenario programs, that are typically encountered in the scenario approach literature.

*A. The fully-supported case*

In this section we assume that, with $\mathbb{P}^m$-probability one, the cardinality of the support set of problem $P_k$, $k = 0, \ldots, \ell$, is equal to $d$. We formalize this in the following assumption.

**Assumption 2** (Fully-supportedness)**.** *For all $k \in \{0, \ldots, \ell\}$, $P_k$ is fully-supported with $\mathbb{P}^m$-probability one.*

Under Assumption 2, we let

$$R_k(S) = \text{supp}(x_k^\star(S)), \quad k \in \{0, \ldots, \ell-1\}, \qquad (18)$$

i.e., we remove the support set of the corresponding optimal solution of $P_k$. Note that the cardinality of $R_k(S)$ is equal to $d$ and this choice for the removed scenarios guarantees that the objective function decreases at each stage, thus improving performance. Moreover, for $k = \ell$, we denote by $R_\ell(S)$ the support set of $x_\ell^\star(S)$. Note that $R_\ell(S)$ does not contain any removed scenarios.

Under Assumption 2, we obtain a tighter the bound than that of Theorem 1, as shown in the following theorem.

**Theorem 3.** *Consider Assumptions 1 and 2. Fix $\epsilon \in (0, 1)$, set $r = \ell d$ and let $m > r + d$. Consider also the scenario discarding scheme as encoded by (18) and illustrated in Figure 2, and let the minimizer of the $\ell$-th program be $x^\star(S) = x_\ell^\star(S)$. We then have that*

$$\mathbb{P}^m\Big\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\big\{\delta \in \Delta : g(x^\star(S), \delta) > 0\big\} > \epsilon\Big\}$$
$$\leq \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \quad (19)$$

The proof of Theorem 3 is deferred to Section V-A. It is important to note that inequality (19) does not involve the combinatorial factor $\binom{r+d-1}{r}$ as in Theorem 1. For a fixed number of scenarios, probability of constraint violation, and confidence level, one is able to satisfy (19) with a larger number of removed scenarios compared to (2). As more scenarios can be discarded using Theorem 3, we can often obtain a better cost improvement. We demonstrate this numerically in Section VII. Besides, note that unlike [25], [26], we do not require the removed scenarios to be violated by the resulting solution (see Figure 3), thus we tighten the bound while relaxing the underlying assumptions.

To illustrate how the proposed scheme works, we consider the pictorial example of Figure 3. Note that $d = 2$, $m = 6$, and we remove $r = 4$, thus requiring $\ell = 2$ steps of the removal scheme of Figure 2. All the problems $P_k$, $k \in \{0, 1, 2\}$, are fully-supported, thus satisfying Assumption 2. The objective function is given by $c^\top x = x_2$ and is indicated by the downwards pointing arrow. The corresponding solution
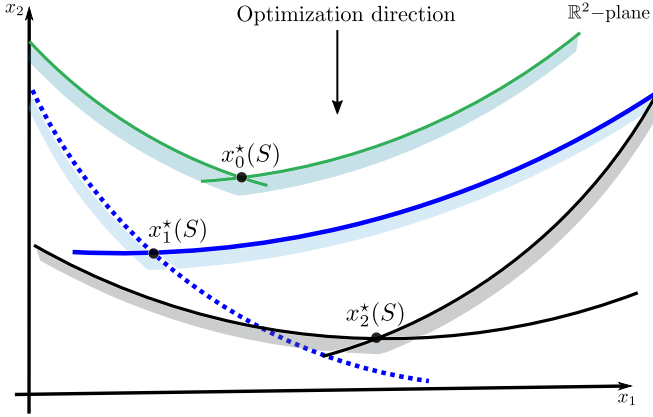
Fig. 3: Pictorial example that illustrates the scheme proposed in Figure 2 for fully-supported problems. In this case, we have that $d = 2, m = 6, r = 4$, and $\ell = 2$, and all the problems $P_k$, $k \in \{0, 1, 2\}$, satisfy Assumption 2. The objective function is given by $c^\top x = x_2$ (indicated by the downwards pointing arrow). The constraint sets are denoted by the different colors patterns: the green constraints are the samples in $\mathrm{supp}(x_0^\star(S))$, the blue ones in $\mathrm{supp}(x_1^\star(S))$, and the black ones in $\mathrm{supp}(x_2^\star(S))$. Note that the dashed-blue constraint is removed by the scheme of Figure 2, but it is not violated by $x_2^\star(S)$.

for the intermediate problems is illustrated by $x_k^\star(S)$, for $k \in \{0, 1, 2\}$, and the support set of each stage by different colour patterns. For instance, the green constraints are the support set, namely, $\mathrm{supp}(x_0^\star(S))$, of problem $P_0$. The shaded colour under each constraint corresponds to the region of the plane that violates that given constraint, e.g., we notice that $x_1^\star(S)$ violates both constraints that belong to $\mathrm{supp}(x_0^\star(S))$ and satisfies all the remaining ones. The result of Theorem 3 provides guarantees for the probability of violation of $x_2^\star(S)$. Note that the dashed-blue constraint is removed at stage 1, but it is not violated by the final solution of our scheme.

### B. The non-degenerate case

In this subsection, we assume that problem $P_k$, $k \in \{0, \ldots, \ell\}$, is non-degenerate.

**Assumption 3** (Non-degeneracy). *For all $k \in \{0, \ldots, \ell\}$, $P_k$ is non-degenerate with $\mathbb{P}^m$-probability one.*

In case of a non fully-supported problem ($\mathrm{supp}(x_k^\star(S)) < d$, for some $k \in \{0, \ldots, \ell\}$), we adopt a procedure called regularization, in the same spirit as in [26]. This is based on introducing a lexicographic order as a tie-break rule to select which additional scenarios to append to $\mathrm{supp}(x_k^\star(S))$, thus constructing a set of cardinality $d$. Note that unless we impose such an order there is no unique choice as all scenarios that are not included in $\mathrm{supp}(x_k^\star(S))$ are not of support, hence their presence leaves the optimal solution unaltered. To this end, we put a unique linear order on the elements of $S$, i.e., assigning them a distinct numerical label. For each $k \in \{0, \ldots, \ell\}$, let
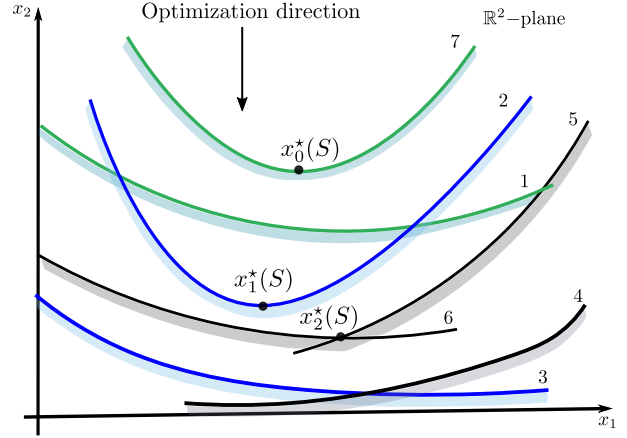


Fig. 4: Illustration of the scheme in Figure 2 applied to non-degenerate, but not fully-supported, problems. The intermediate solutions are denoted by $x_k^\star(S)$, $k = 0, 1, 2$. The different colour patterns depict the removed scenarios at each stage. The green constraints are the ones removed at stage 0, the blue ones those removed at stage 1. Similar as before, the objective function is given by $c^\top x = x_2$ and this is indicated by the downwards arrow. Observe that the optimal solution, consequently the final solution returned by our scheme, depends on the linear order imposed to the original scenarios.

$\nu_k(S) = d - |\mathrm{supp}(x_k^\star(S))|$ and define the following sets recursively as

$$Z_k(S) = \Bigg\{ \nu_k(S) \text{ scenarios with the smallest labels in}$$

$$S \setminus \Bigg( \bigcup_{j=0}^{k-1} \big\{ \mathrm{supp}(x_j^\star(S)) \cup Z_j(S) \big\} \cup \mathrm{supp}(x_k^\star(S)) \Bigg) \Bigg\}, \quad (20)$$

with $Z_0(S)$ containing the $\nu_0(S)$ smallest elements of $S \setminus \mathrm{supp}(x_0^\star(S))$ according to the linear order. Note that the set appearing in the the definition of $Z_k(S)$ in (20) corresponds to scenarios available at stage $k$ that are not of support.

For each $k \in \{0, \ldots, \ell - 1\}$, we can now define the sets of discarded samples as

$$R_k(S) = \mathrm{supp}(x_k^\star(S)) \cup Z_k(S). \quad (21)$$

Notice that by construction $|R_k(S)| = d$, while if for any $k \in \{0, \ldots, \ell\}$, $P_k$ is fully-supported, then $R_k(S) = \mathrm{supp}(x_k^\star(S))$, i.e., it coincides with the support set of $x_k^\star(S)$. Similar as in the fully-supported case, we denote by $R_\ell(S)$ the superset of the support set of $x_\ell^\star(S)$ obtained by appending, if necessary, $\nu_\ell(S)$ scenarios from the remaining ones.

Notice that similarly to the fully-supported case, the objective function will necessarily decrease at each stage and we do not require that all scenarios discarded up to stage $k - 1$, i.e., $\bigcup_{j=0}^{k-1} R_j(S)$, are violated by $x_k^\star(S)$.

**Remark 1.** *Consider two arbitrary scenario sets $C \subset C'$, and denote by $x_k^\star(C)$ and $x_k^\star(C')$ the minimizers of $P_k$ with $C$ and $C'$, respectively, replacing $S$. Moreover, define $Z_k(C)$ and $Z_k(C')$ as in (20) with $C$ and $C'$, respectively, in place*

of $S$. We then have that $F_k(x_k^\star(C)) < F_k(x_k^\star(C'))$ if: either $c^\top x_k^\star(C) < c^\top x_k^\star(C')$; or $c^\top x_k^\star(C) = c^\top x_k^\star(C')$ and, at the first element that $Z_k(C)$ and $Z_k(C')$ differ, the corresponding label of $Z_k(C)$ is strictly lower with respect to the imposed lexicographic order than the one of $Z_k(C')$. Regularization is thus a way to select among subsets of scenarios that would otherwise yield the same objective value. We will use this procedure in Section V-B to prove Theorem 4 below. It is shown in [26], that $P_k$ with its objective function replaced by

$$F_k(x) = \big(c^\top x, Z_k(S)\big). \tag{22}$$

is a fully-supported program, and the constructed set $R_k(S)$ in (21) forms its unique support set with cardinality $d$.

We are now in position to state the main result related to non-degenerate problems.

**Theorem 4.** *Consider Assumptions 1 and 3. Fix $\epsilon \in (0,1)$, set $r = \ell d$ and let $m > r + d$. Consider also the scenario discarding scheme as encoded by (21) and illustrated in Figure 2, and let the minimizer of the $\ell$-th program be $x^\star(S) = x_\ell^\star(S)$. We then have that*

$$\mathbb{P}^m\Big\{(\delta_1,\ldots,\delta_m) \in \Delta^m : \mathbb{P}\big\{\delta \in \Delta : g(x^\star(S),\delta) > 0\big\} > \epsilon\Big\}$$
$$\leq \sum_{i=0}^{r+d-1} \binom{m}{i}\epsilon^i(1-\epsilon)^{m-i}. \tag{23}$$

Theorem 4 holds for non-degenerate scenario programs, thus being more general than Theorem 3, which is only valid for fully-supported problems. This generalization comes at the expense of a (possible) decrease in performance, as we append additional scenarios to compose the support set of the regularized problem that may not improve the objective value associated to $x^\star(S)$. However, similar to Theorem 3 we may still improve the cost with respect to other removal strategies, as we are allowed to remove more constraints compared to [25]. It is also important to notice that (23) holds for any linear order imposed in the original samples $S$ and that the resulting optimal objective value of the scheme depends on such ordering. In this paper we only provide feasibility guarantees and not optimality. Note that this is also the case for the results in [25], [26]. The only available results for the optimal cost are given in [25] when the removal scheme is the optimal one, which is, however, of combinatorial complexity. As a final remark, observe that Theorem 3 constitutes the main development towards obtaining Theorem 4, as should be apparent in the next section.

**Remark 2.** *It should be noted that the assumption in [25], [26] appearing in the statement of Theorem 1, that requires all discarded scenarios to be violated by the final solution with $\mathbb{P}^m$-probability one, has some non-degeneracy implications for all intermediate problems. To see this, notice that if we allow for degenerate problems, then pathological situations where all scenarios are identical are admissible and may happen with non-zero probability (e.g., allowing for atomic masses). Clearly, in such cases there is no scenario that can* be discarded while being violated by the resulting solution which remains unaltered. Therefore, we tighten the bound in Theorem 4, without strengthening the assumptions in [25], [26].

To clarify how the scheme presented in Figure 2 works when applied to non-degenerate problems, consider the example depicted in Figure 4. Similar as before, we have $d = 2, m = 7$, and want to remove 4 constraints, i.e., $r = 4$. As opposed to Figure 3, however, note that the constraints are enumerated according to an arbitrary order, which is used to compose the sets $Z_k(S)$, $k \in \{0,1,2\}$, as described by equation (20). Moreover, problems $P_0$ and $P_1$ are not fully-supported, as the number of support scenarios is equal to one in each of these cases. Our scheme first removes the scenario that supports the solution $x_0^\star(S)$ and the one labeled as 1, since it is the scenario with the smallest order among the remaining ones. These scenarios are depicted as green in Figure 4. Then, we solve problem $P_1$ with the resulting scenarios, obtaining $x_1^\star(S)$ as an intermediate solution and scenarios labeled as 2 and 3 to be removed. The former constraint is removed as it is in the support set of $x_1^\star(S)$, and the latter as it is the sample with the smallest index from the remaining ones. Finally, the solution provided by the scheme, and whose guarantees are given in Theorem 4, is denoted by $x_2^\star(S)$.

## V. PROOF OF THE MAIN RESULTS

### A. The fully-supported case

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

$$z^\star(I) = \operatorname*{argmin}_{x \in X} \quad c^\top x$$
$$\text{subject to} \quad g(x,\delta) \leq 0, \text{ for all } \delta \in I. \tag{24}$$

Recall that $x_k^\star(C)$ denotes the minimizer of $P_k$ which in turn is based on the samples in $C \setminus \cup_{j=0}^{k-1} R_j(C)$, i.e., the ones that have not been removed up to stage $k$ of the proposed scheme. It thus holds that $x_k^\star(C) = z^\star(C \setminus \cup_{j=0}^{k-1} R_j(C))$ – note that the argument of $z^\star$ in this case depends on $k$, $k \in \{0,\ldots,\ell\}$. Recall also that, under Assumption 2, we have $R_k(C) = \operatorname{supp}(x_k^\star(C))$. Since we will be invoking the framework introduced in Section II-B, let

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

and define the mapping $\mathcal{A} : \Delta^m \to 2^\Delta$, with $\zeta = (\ell+1)d$, as

$$\mathcal{A}(C) = \Big(\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\Big) \cup \mathcal{A}_3(C). \tag{25}$$

The main motivation to define the mapping in (25) is the fact that its probability of violation will be shown to upper bound that of $\{\delta \in \Delta : g(x_\ell^\star(C), \delta) \leq 0\}$, which is ultimately the quantity we are interested in (as shown in Section V-A, step 3).

Note that $\mathcal{A}(C)$ comprises three sets:

i) The set $\mathcal{A}_1(C)$ contains all realizations of $\delta$ for which the final decision of our proposed scheme $x_\ell^\star(C) = x^\star(C)$ remains feasible. This is the set whose probability of occurrence we are ultimately interested to bound.

ii) The set $\mathcal{A}_2(C)$, which is composed by the intersection of $\ell + 1$ sets indexed by $k \in \{0, \dots, \ell\}$, contains the realizations of $\delta$ such that, for all subsets of cardinality $d - 1$ from the remaining samples at stage $k$, the cost $c^\top z^\star(J \cup \{\delta\})$ is lower than or equal to $c^\top x_k^\star(C)$. The former cost corresponds to appending $\delta$ to any set $J$ of $d - 1$ scenarios from $C \setminus \cup_{j=0}^{k-1} R_j(C)$, while the latter corresponds to the cost of the minimizer $x_k^\star(C)$ of $P_k$. Informally, this inequality is of similar nature with that of $\mathcal{A}_1(C)$, however, rather than considering constraint satisfaction it only involves some cost dominance condition for each of the interim and the final optimal solutions. The motivation to use this representation rather than constraint satisfaction conditions stems from the fact that in Section IV-B we will be appending a lexicographic order to the cost so that we break the tie among multiple compression sets. Besides, these sets carry information about the path taken by the proposed scheme, which is to be understood, in this context, as the sequence $(x_k(C))_{k=0}^\ell$.

iii) The set $\mathcal{A}_3(C)$ includes all scenarios that are removed by the discarding scheme. Implicit in the definition of mapping (25) is the fact that, for any compression set $C$, all samples that are not removed in the intermediate stages must be contained in the set $\mathcal{A}_1(C) \cap \mathcal{A}_2(C)$. This fact will be crucial in the following arguments.

The following proposition establishes a basic property of any compression associated to the mapping (25), and is instrumental for the proof of our main theorem.

**Proposition 1.** *Consider Assumptions 1 and 2. Set $r = \ell d$ and let $m > (\ell + 1)d$. We have that $C \subset S$ is a compression set for $\mathcal{A}(C)$ in (25) if and only if, for all $k \in \{0, \dots, \ell\}$,*

$$x_k^\star(C) = x_k^\star(S). \tag{26}$$

*Proof.* We first show necessity. Suppose that $C$ is a compression set but, for the sake of contradiction, we have that there exists $k \in \{0, \dots, \ell\}$ and $\bar{\delta} \in S \setminus C$ such that

$$x_k^\star(C) \neq x_k^\star(C \cup \{\bar{\delta}\}). \tag{27}$$

Let $\bar{k}$ be the minimum index such that (27) holds, while we have that $x_j^\star(C) = x_j^\star(C \cup \{\bar{\delta}\})$, for all $j < \bar{k}$.

By Assumption 2, the last statement implies that $\text{supp}(x_j^\star(C)) = \text{supp}(x_j^\star(C \cup \{\bar{\delta}\}))$, for all $j < \bar{k}$, as the support set of each optimal solution is unique. Hence, $R_j(C) = R_j(C \cup \{\bar{\delta}\})$ for all $j < \bar{k}$, and

$R_j(C) = \text{supp}(x_j^\star(C))$ for fully-supported problems (similarly for $R_j(C \cup \{\bar{\delta}\})$). By (24), we then have

$$x_{\bar{k}}^\star(C) = z^\star(C \setminus \cup_{j=0}^{\bar{k}-1} R_j(C)), \tag{28}$$

$$x_{\bar{k}}^\star(C \cup \{\bar{\delta}\}) = z^\star((C \setminus \cup_{j=0}^{\bar{k}-1} R_j(C)) \cup \{\bar{\delta}\}). \tag{29}$$

Since the right-hand side of (29) involves one more scenario with respect to the right-hand side of (28), the feasible set of (29) is a subset set of the one of (28). Moreover, by the fact that $x_{\bar{k}}^\star(C \cup \{\bar{\delta}\}) \neq x_{\bar{k}}^\star(C)$ and Assumption 1, we obtain that

$$c^\top x_{\bar{k}}^\star(C) < c^\top x_{\bar{k}}^\star(C \cup \{\bar{\delta}\}). \tag{30}$$

Notice that $\bar{\delta}$ belongs to the support set of $x_{\bar{k}}^\star(C \cup \{\bar{\delta}\})$, as its removal results in a different optimal solution with lower cost in (30). In other words, there exists $\bar{J} \subset C \setminus \cup_{j=0}^{\bar{k}-1} R_j(C)$ (in fact, $\bar{J} = \text{supp}(x_{\bar{k}}^\star(C \cup \{\bar{\delta}\})) \setminus \{\bar{\delta}\}$) of cardinality $d - 1$ such that by (29), we have that

$$c^\top x_{\bar{k}}^\star(C) < c^\top x_{\bar{k}}^\star(C \cup \{\bar{\delta}\}) = c^\top z^\star(\bar{J} \cup \{\bar{\delta}\}). \tag{31}$$

At the same time, $C$ is assumed to be a compression set. Since $\bar{\delta} \notin C$, then $\bar{\delta} \notin \cup_{k=0}^{\ell-1} R_k(C) = \mathcal{A}_3(C)$, as $\cup_{k=0}^{\ell-1} R_k(C) \subset C$. As a result, $\bar{\delta}$ will give rise to a constraint in $P_\ell$, hence $\bar{\delta} \in \mathcal{A}_2(C)$, which in turn implies that for all $J \subset C \setminus \cup_{j=0}^{\ell-1} R_j(C)$ with $|J| = d - 1$, and for all $k \leq \ell$,

$$c^\top z^\star(J \cup \{\bar{\delta}\}) \leq c^\top x^\star(C) \leq c^\top x_k^\star(C), \tag{32}$$

where the first inequality follows from the fact that $c^\top x^\star(C)$ is the optimal value for $P_\ell$, and $x^\star(C) = x_\ell^\star(C)$ by construction satisfies all constraints with scenarios in $J \cup \{\bar{\delta}\}$. The second inequality follows from the fact that $k \leq \ell$, and the cost deteriorates as $k$ increases. Setting $k = \bar{k}$ and $J = \bar{J}$ in (32) establishes a contradiction with (31), thus showing that $x_k^\star(C) = x_k^\star(C \cup \{\delta\})$, for any $\delta \in S \setminus C$, and any $k \in \{0, \dots, \ell\}$. Proceeding inductively, adding one by one each element in $S \setminus C$, we can show that $x_k^\star(C) = x_k^\star(S)$, for any $k \in \{0, \dots, \ell\}$, thus concluding the necessity part of the proof.

We now show sufficiency. Let $C \subset S$ be such that $x_k^\star(C) = x_k^\star(S)$ for all $k \in \{0, \dots, \ell\}$. We aim to show that $C$ is a compression for $S$, i.e., $\delta \in \mathcal{A}(C)$ for all $\delta \in S$. Recalling the definition of the mapping $\mathcal{A}(C)$ from (25) we note that, under this scenario, the sets $\mathcal{A}_1(C)$ and $\mathcal{A}_3(C)$ are trivially equal to $\mathcal{A}_1(S)$ and $\mathcal{A}_3(S)$, respectively. Moreover, since $C \subset S$ and $x_k^\star(C) = x_k^\star(S)$ for all $k \in \{0, \dots, \ell\}$, which implies that $R_k(C) = R_k(S)$ by Assumption 1, we have that $S \setminus \cup_{j=0}^{k-1} R_j(S) = S \setminus \cup_{j=0}^{k-1} R_j(C) \supset C \setminus \cup_{j=0}^{k-1} R_j(C)$. The latter implies then that the inequalities in $\mathcal{A}_2(S)$ constitute a superset of those in $\mathcal{A}_2(C)$, hence, that problem is more constrained and as a result $\mathcal{A}_2(S) \subset \mathcal{A}_2(C)$. By construction we have that $\delta \in \mathcal{A}(S)$ for all $\delta \in S$. This in turn implies that if a sample is not removed, then it will have to be included in $\mathcal{A}_2(S)$, and due to the established inclusion also in $\mathcal{A}_2(C)$. Since $\mathcal{A}_1(S) = \mathcal{A}_1(C)$ and $\mathcal{A}_3(S) = \mathcal{A}_3(C)$, we then have that $\delta \in \mathcal{A}(C)$ for all $\delta \in S$, showing that $C$ is a compression set. This concludes the proof of the proposition. $\square$

*Proof of Theorem 3:* A natural compression candidate is

$$C = \bigcup_{k=0}^{\ell} \text{supp}(x_k^\star(S)), \qquad (33)$$

as it consists of the support sets of the intermediate problems.

*Existence:* We prove that $C$ in (33) is a compression set. By the sufficiency part of Proposition 1, it suffices to show that the set $C$ in (33) satisfies $x_k^\star(C) = x_k^\star(S)$, for all $k \in \{0, \ldots, \ell\}$. We will show this by means of induction. For the base case $k = 0$, notice that

$$c^\top x_0^\star(S) = c^\top z^\star(S) = c^\top z^\star(\text{supp}(x_0^\star(S)))$$
$$= c^\top x_0^\star(C), \qquad (34)$$

where the first equality is due to (24), the second equality is due to the fact that $\text{supp}(x_0^\star(S))$ is the support set of $x_0^\star(S)$, while the last equality is due to Assumption 2, the definition of support set and the fact that $\text{supp}(x_0^\star(S)) \subset C$. By (34), and Assumption 1, we conclude that $x_0^\star(C) = x_0^\star(S)$.

To complete the induction argument, we assume that $x_j^\star(C) = x_j^\star(S)$ for all $j \in \{0, \ldots, \bar{k}\}$, for some $\bar{k} < \ell$. We will show that $x_{\bar{k}+1}^\star(C) = x_{\bar{k}+1}^\star(S)$. To this end, by Assumption 2, $x_j^\star(C) = x_j^\star(S)$ for all $j \le \bar{k}$ implies that $\text{supp}(x_j^\star(C)) = \text{supp}(x_j^\star(S))$, for all $j \le \bar{k}$, as the support set of each optimal solution is unique. Moreover, $R_j(C) = R_j(S)$ for all $j < \bar{k}$, as $R_j(C) = \text{supp}(x_j^\star(C))$ for fully-supported problems. Similarly to the base case we have that

$$c^\top x_{\bar{k}+1}^\star(C) = c^\top z^\star(C \setminus \cup_{j=0}^{\bar{k}} R_j(S))$$
$$\le c^\top z^\star(S \setminus \cup_{j=0}^{\bar{k}} R_j(S)) = c^\top x_{\bar{k}+1}^\star(S), \quad (35)$$

where the first and last equalities are due to (24), and the inequality is due to the fact that

$$C \setminus \cup_{j=0}^{\bar{k}} R_j(S) \subseteq S \setminus \cup_{j=0}^{\bar{k}} R_j(S).$$

Moreover

$$c^\top x_{\bar{k}+1}^\star(S) = c^\top z^\star(S \setminus \cup_{j=0}^{\bar{k}} R_j(S))$$
$$= c^\top z^\star(\text{supp}(x_{\bar{k}+1}^\star(S))) \le c^\top z^\star(C \setminus \cup_{j=0}^{\bar{k}} R_j(S))$$
$$= c^\top x_{\bar{k}+1}^\star(C), \qquad (36)$$

where the first and last equalities are due to (24), the second one due to the fact that $\text{supp}(x_{\bar{k}+1}^\star(S)) \subset S \setminus \cup_{j=0}^{\bar{k}} R_j(S)$, and the inequality holds since $R_j(C) = R_j(S)$ and $\text{supp}(x_{\bar{k}+1}^\star(S)) \subset C \setminus \cup_{j=0}^{\bar{k}} R_j(S)$. By (35) and (36) we then have that $x_{\bar{k}+1}^\star(C) = x_{\bar{k}+1}^\star(S)$, thus concluding the induction proof. In other words, we have shown that

$$x_k^\star(C) = x_k^\star(S), \text{ for all } k \in \{0, \ldots, \ell\}. \qquad (37)$$

Relation (37) together with the sufficiency part of Proposition 1 shows that the candidate $C$ in (33) is a compression set.

*Uniqueness:* To show that $C$ in (33) is the unique compression set, assume for the sake of contradiction that there exists another compression $C' \subset S$ for the mapping defined in (25), $C' \ne C$, with $|C'| = (\ell+1)d$. Since $C' \subset S$ is a compression,

Proposition 1 (necessity part) implies that $x_k^\star(C') = x_k^\star(S)$, for all $k \in \{0, \ldots, \ell\}$, as $C'$ is a compression. Besides, by the existence part (Step 1 above), we have shown that for $C$ given in (33) we have that $x_k^\star(C) = x_k^\star(S)$ for all $k \in \{0, \ldots, \ell\}$. We thus have that for all $k \in \{0, \ldots, \ell\}$, $x_k^\star(C) = x_k^\star(C')$. This in turn implies that $\text{supp}(x_k^\star(C)) = \text{supp}(x_k^\star(C'))$ for all $k \in \{0, \ldots, \ell\}$, which, by Assumption 2, leads to $C = C'$ (to see this notice that $\cup_{k=0}^{\ell}\text{supp}(x_k^\star(S)) \subset C'$ and $|C'| = (\ell+1)d$), thus establishing a contradiction.

*Linking Theorem 2 with the probability of constraint violation:* Recall that

$$\mathcal{A}(C) = \Big(\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\Big) \cup \mathcal{A}_3(C), \qquad (38)$$

where the individual sets are as in (25). Recall also that $\mathcal{A}_3(S)$ is a discrete set that contains the removed samples throughout the execution of the scheme of Figure 2. Fix any $S$ with $m$ scenarios, set $r = \ell d$ and let $m > (\ell+1)d$. Fix also $\epsilon \in (0,1)$. Let $C \subset S$ with $|C| = (\ell+1)d$ be the unique compression defined in (33). We have that

$$\mathbb{P}\{\mathcal{A}(C)\} = \mathbb{P}\{(\mathcal{A}_1(C) \cap \mathcal{A}_2(C)) \cup \mathcal{A}_3(C)\}$$
$$= \mathbb{P}\{\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\},$$
$$\le \mathbb{P}\{\mathcal{A}_1(C)\} = \mathbb{P}\{\delta \in \Delta : g(x^\star(C), \delta) \le 0\},$$
$$= \mathbb{P}\{\delta \in \Delta : g(x^\star(S), \delta) \le 0\}, \qquad (39)$$

where the first equality is due to the fact that $\mathbb{P}\{\mathcal{A}_3(C)\} = 0$, since $\mathcal{A}_3(C)$ is a discrete set and we have imposed the non-degeneracy condition of Assumption 3 which prevents scenarios to have accumulation points with non-zero probability, while the inequality is due to the fact that $\mathcal{A}_1(C) \cap \mathcal{A}_2(C) \subseteq \mathcal{A}_1(C)$. The second last equality is by definition of $\mathcal{A}_1(C)$, and the last one follows from the fact that $x^\star(C) = x^\star(S)$ (see (37)).

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

$$\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^\star(S), \delta) > 0\} > \epsilon\}$$
$$\le \mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \notin \mathcal{A}(C)\} > \epsilon\}. \qquad (40)$$

Therefore, since set $C$ in (33) is the unique compression of $\mathcal{A}(C)$, by Theorem 2 we have that

$$\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : \delta \notin \mathcal{A}(C)\} > \epsilon\}$$
$$\le \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \qquad (41)$$

By (40) and (41) we then have that $\mathbb{P}^m\{(\delta_1, \ldots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^\star(S), \delta) > 0\} > \epsilon\} \le \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$, thus concluding the proof of Theorem 3. $\square$

### B. The non-degenerate case

Throughout this subsection, we consider Assumption 3. Let $m > (\ell+1)d$, and consider any set $C \subset S$ with $|C| = (\ell+1)d$.

We modify the mapping $\mathcal{A}(C)$ in (25) by replacing the second set in its definition with

$$\mathcal{A}_2(C) = \bigcap_{k=0}^{\ell} \big\{ \delta \in \Delta : F_k(z^\star(J \cup \{\delta\})) \leq F_k(x_k^\star(C)),$$
$$\text{for all } J \subset C \setminus \cup_{j=0}^{k-1} R_j(C), \text{ with } |J| = d - 1 \big\}, \quad (42)$$

where $F_k(\cdot)$ is the augmented objective function defined in (22), related to $P_k$ defined by means of the regularization procedure of Section IV. The above inequality is to be understood in a lexicographic sense as detailed in Remark 1. A natural candidate compression set in this case is

$$C = \bigcup_{k=0}^{\ell} (\text{supp}(x_k^\star(S)) \cup Z_k(S)), \quad (43)$$

which is composed by the removed samples of the scheme, and the support set of the last stage together with the corresponding constraints in $Z_\ell(S)$. In fact, we now append $Z_k(S)$ in the definition of $C$ to ensure that $|C| = (\ell+1)d$, as $|\text{supp}(x_k^\star(S))|$ could be lower than $d$ as the intermediate problems might not be fully-supported. Similarly to the fully-supported case, our goal is to show that the compression set defined in (43) is the unique compression set of size $(\ell+1)d$ for the mapping in (25), with $\mathcal{A}_2(C)$ in (42) in place of $\mathcal{A}_2(C)$ in (25). By (21), recall that $R_k(C) = \text{supp}(x_k^\star(C)) \cup Z_k(C)$, $k \in \{0, \dots, \ell-1\}$.

**Proposition 2.** *Suppose Assumptions 1 and 3 hold. Let $C$ be the set in* (43)*, and consider the scheme of Figure 2 with the removed scenarios given by* (21)*. We have that:*

i) $x_k^\star(C) = x_k^\star(S)$ *and* $Z_k(C) = Z_k(S)$ *for all* $k \in \{0, \dots, \ell\}$.

ii) *Let $C'$ be any other compression of size $(\ell+1)d$. Suppose $R_j(C) = R_j(C')$ for all $j \in \{0, \dots, \bar{k}-1\}$, where $\bar{k}$ is the smallest index such that $x_{\bar{k}}^\star(C') \neq x_{\bar{k}}^\star(C)$. Then $x_{\bar{k}}^\star(C') \neq x_{\bar{k}}^\star(C' \cup \{\delta\})$ for some $\delta \in \text{supp}(x_{\bar{k}}^\star(C)) \setminus \text{supp}(x_{\bar{k}}^\star(C'))$. Moreover, such a $\delta$ is in fact in the set $C \setminus C'$.*

*Proof. Item i):* We use induction. Fix $k = 0$ and note that

$$x_0^\star(C) = z^\star(C) = z^\star(\text{supp}(x_0^\star(S))) = x_0^\star(S), \quad (44)$$

where the first equality follows from the definition in (24), for the second one we use the definition of the support set, and the third one follows from the definition of $x_0^\star(S)$ and the definition of the support set. Moreover, we have that

$$Z_0(C) = \big\{ \nu_0(S) \text{ scenarios with the smallest labels in}$$
$$C \setminus \{\text{supp}(x_0^\star(S))\} \big\}$$
$$= \big\{ \nu_0(S) \text{ scenarios with the smallest labels in}$$
$$S \setminus \{\text{supp}(x_0^\star(S))\} \big\} = Z_0(S), \quad (45)$$

where the first equality is due to the definition of $C$ in (43) and the fact that $Z_0(S) \subset C$, while the last one is due to the definition of $Z_0(S)$ in (20). Assume now that $x_k^\star(C) = x_k^\star(S)$

and $Z_k^\star(C) = Z_k^\star(S)$ for all $k \in \{0, \dots, \bar{k}\}$, and consider the case $\bar{k} + 1$. Indeed, we have that

$$x_{\bar{k}+1}^\star(C) = z^\star(C \setminus \cup_{j=0}^{\bar{k}} R_j(C)) = z^\star(\text{supp}(x_{\bar{k}+1}^\star(S)))$$
$$= z^\star(S \setminus \cup_{j=0}^{\bar{k}} R_j(S)) = x_{\bar{k}+1}^\star(S), \quad (46)$$

where these relations follow as in (44) for the case $k = 0$. We also have that

$$Z_{\bar{k}+1}^\star(C) = \Big\{ \nu_{\bar{k}+1}(S) \text{ scenarios with the smallest labels in}$$
$$C \setminus \Big( \bigcup_{j=0}^{\bar{k}} R_j(S) \cup \text{supp}(x_{\bar{k}+1}^\star(S)) \Big) \Big\} = Z_{\bar{k}+1}^\star(S), \quad (47)$$

since $Z_{\bar{k}+1}^\star(S) \subset C \setminus \bigcup_{j=0}^{\bar{k}} \{R_j(S) \cup \text{supp}(x_{\bar{k}+1}^\star(S))\}$ due to the particular choice of $C$ in (43), thus proving that for $C$ in (43) we have $x_k^\star(C) = x_k^\star(S)$ and $Z_k(C) = Z_k(S)$, for all $k \in \{0, \dots, \ell\}$. This concludes the proof of item $i$).

*Item ii):* We prove the contrapositive. Assume that for all $\delta \in \text{supp}(x_{\bar{k}}^\star(C)) \setminus \text{supp}(x_{\bar{k}}^\star(C'))$ we have that $x_{\bar{k}}^\star(C') = x_{\bar{k}}^\star(C' \cup \{\delta\})$. We will show that $x_{\bar{k}}^\star(C) = x_{\bar{k}}^\star(C')$. We then have that

$$c^\top x_{\bar{k}}^\star(C') = c^\top x_{\bar{k}}^\star(C' \cup \{\delta\}) = c^\top x_{\bar{k}}^\star(C' \cup \text{supp}(x_{\bar{k}}^\star(C)))$$
$$= c^\top x_{\bar{k}}^\star(C), \quad (48)$$

where the second equality holds due to Lemma 2.12 in [26] since $C' \cup \{\delta\} \subset C' \cup \text{supp}(x_{\bar{k}}^\star(C))$. The last equality follows from the definition of the support set and the non-degeneracy condition of Assumption 3. By Assumption 1 we then conclude that $x_{\bar{k}}^\star(C) = x_{\bar{k}}^\star(C')$.

We now show that such a $\delta$ must belong to $C \setminus C'$. In fact, choose $\bar{\delta} \in \text{supp}(x_{\bar{k}}^\star(C)) \setminus \text{supp}(x_{\bar{k}}^\star(C'))$ and assume for the sake of contradiction that $\bar{\delta} \in C'$. This implies that $\bar{\delta} \in R_j(C')$ for some $j \geq \bar{k}$. In this is the case, we have that

$$c^\top x_{\bar{k}}^\star(C' \cup \{\bar{\delta}\}) = c^\top z^\star \Big( \Big(C' \setminus \bigcup_{j=0}^{\bar{k}-1} R_j(C)\Big) \cup \{\bar{\delta}\} \Big)$$
$$= c^\top z^\star (\text{supp}(x_{\bar{k}}^\star(C')))$$
$$= c^\top x_{\bar{k}}^\star(C') \quad (49)$$

where the first relation holds due to (24) and the fact that $R_j(C') = R_j(C)$ for all $j < \bar{k}$, the second one is due to the fact that $\text{supp}(x_{\bar{k}}^\star(C')) \subset C' \setminus \bigcup_{j=0}^{\bar{k}-1} R_j(C) \cup \{\bar{\delta}\}$ and $\bar{\delta} \in R_j(C')$ for $j \geq \bar{k}$. The third equality follows from the definition of the support set and the non-degeneracy condition of Assumption 3. However, note that (49) contradicts our choice of $\bar{\delta}$, which requires that $x_{\bar{k}}^\star(C') \neq x_{\bar{k}}^\star(C' \cup \{\bar{\delta}\})$. This concludes the proof. $\square$

*Proof of Theorem 4: Existence.* The existence part follows *mutatis mutandis* from the one of Theorem 3. In fact, $\mathcal{A}_1(C) = \mathcal{A}_1(S)$ and $\mathcal{A}_3(C) = \mathcal{A}_3(S)$ by Proposition 2, item $i$), and $\mathcal{A}_2(S) \subset \mathcal{A}_2(C)$ as $C \subset S$ (see the discussion at the end of Proposition 1). *Uniqueness:* Let $C'$ be another compression of size $(\ell+1)d$ and assume for the sake of contradiction that $C \neq C'$. We can distinguish two possible cases. Case I: there exists a $\bar{k} \in \{0, \dots, \ell\}$ such that $x_{\bar{k}}^\star(C') \neq x_{\bar{k}}^\star(C)$; or case

II: $x_k^\star(C') = x_k^\star(C)$ for all $k \in \{0, \ldots, \ell\}$, but there exists a $\tilde{k} \in \{0, \ldots, \ell\}$ such that $Z_{\tilde{k}}(C') \neq Z_{\tilde{k}}(C)$. In the sequel, we argue separately that neither of these cases can happen.

*Case I:* Let $\bar{k}$ be the smallest index such that $x_{\bar{k}}^\star(C') \neq x_{\bar{k}}^\star(C)$, and let $\tilde{k} \leq \bar{k}$ be the smallest index such that $Z_{\tilde{k}}(C') \neq Z_{\tilde{k}}(C)$. Consider first the case where $\tilde{k} < \bar{k}$. Under these definitions, note that $R_j(C') = R_j(C)$ for all $j < \tilde{k}$. Moreover, we have that

$$Z_{\tilde{k}}(C') = \Big\{ \nu_{\tilde{k}}(C) \text{ scenarios with the smallest labels in }$$

$$C' \setminus \bigcup_{j=0}^{\tilde{k}-1} R_j(C') \Big\}$$

$$= \Big\{ \nu_{\tilde{k}}(C) \text{ scenarios with the smallest labels in }$$

$$C' \setminus \bigcup_{j=0}^{\tilde{k}-1} R_j(S) \Big\}, \tag{50}$$

where the first equality is by the definition in (20) and the fact that $\nu_{\tilde{k}}(C') = \nu_{\tilde{k}}(C)$ since $\tilde{k} < \bar{k}$; the second equality follows since $R_j(C') = R_j(C) = R_j(S)$ – the last equality follows from Proposition 2, item $i$) –for all $j \leq \tilde{k} - 1$ and due to the uniqueness requirement of Assumption 1. Note that $Z_{\tilde{k}}(C') \neq Z_{\tilde{k}}(S)$ and $C' \subset S$ implies that, for all $\delta \in Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(S)$,

$$y_\delta > \max_{\xi \in Z_{\tilde{k}}(S)} y_\xi = y_{\max}, \tag{51}$$

where $y_\delta \in \mathbb{N}$ corresponds to the label associated to $\delta$.

We will use the relation (51) to show that any element in $C \setminus C'$ has a label greater than $y_{\max}$. In fact, note that

$$C' \setminus C \subset \Big\{ \cup_{j=\tilde{k}+1}^\ell R_j(C') \Big\} \cup \{ Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(C) \}, \tag{52}$$

hence it suffices to show that any element in either set in the right-hand side of (52) is greater than $y_{\max}$. To this end, fix any $\delta \in \cup_{j=\tilde{k}+1}^\ell R_j(C')$ and note that

$$y_\delta > \max_{\xi \in Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(C)} y_\xi > y_{\max}, \tag{53}$$

where the first inequality is due to the fact that since such a $\delta$ has not been removed up to stage $\tilde{k}$, then its label will be greater than the ones in $Z_{\tilde{k}}(C')$, and as a result the ones in $Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(C)$. The second inequality follows from (51) and the fact that $Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(C) \subset Z_{\tilde{k}}(S)$. Therefore, for any $\delta \in C' \setminus C$ we have that $y_\delta > y_{\max}$.

From now on, let $\delta$ be the scenario associated to $y_{\max}$. Pick $\bar{J} = \{\text{supp}(x_{\tilde{k}}^\star(C))\} \cup \{Z_{\tilde{k}}(C) \setminus \{\delta\}\}$, which has cardinality $d - 1$ and is a subset of $C \setminus \cup_{j=0}^{\tilde{k}-1} R_j(C)$, and fix $\bar{\delta} \in C' \setminus C$. Note that under this choice of $\bar{\delta}$

$$F_{\tilde{k}}(z^\star(\bar{J} \cup \{\bar{\delta}\})) > F_{\tilde{k}}(x_{\tilde{k}}^\star(C)), \tag{54}$$

since $y_{\bar{\delta}} > y_{\max}$ (by our previous discussion) and the inequality is interpreted lexicographically. However, this contradicts the fact that $C$ is a compression set (see Definition 1) as $\bar{\delta} \in C' \setminus C \subset S$, hence $\bar{\delta} \notin \mathcal{A}_3(C')$ has not been removed, but $\bar{\delta} \notin \mathcal{A}_2(C)$ due to (54).

Consider now the case $\tilde{k} = \bar{k}$. Note that, in this case, we have that $R_j(C') = R_j(C)$ for all $j \leq \bar{k} - 1$. Based on the result of Proposition 2, item $ii$), applied to $C'$ (note that the assumptions of Proposition 2, item $ii$), are satisfied with our choice of $C'$), we observe that there exists a $\bar{\delta} \in \{\text{supp}(x_{\bar{k}}^\star(C)) \setminus \text{supp}(x_{\bar{k}}^\star(C'))\} \cap \{C \setminus C'\}$ such that $x_{\bar{k}}^\star(C') \neq x_{\bar{k}}^\star(C' \cup \{\bar{\delta}\})$. Repeating the arguments following equations (28) and (29) in the necessity proof of Proposition 1 with $C'$ in the place of $C$ in that proposition, we reach a contradiction that $C'$ is a compression set.

*Case II:* We can reach a contradiction if case II holds in a similar fashion as in case I. In fact, letting $\tilde{k}$ be the smallest index such that $Z_{\tilde{k}}(C') \neq Z_{\tilde{k}}(C)$, the proof proceeds in an identical manner with case I.

Hence, we conclude that in any case $C = C'$, thus proving uniqueness of the compression set in (43).

*Linking Theorem 2 with the probability of violation:* Note that for the non-degenerate case the mapping has the same structure as the one in (25), with the set $\mathcal{A}_2(C)$ in (25) being substituted with the one in (42). The arguments then follows *mutatis mutandis* the ones in the last part of the fully-supported case. This concludes the proof of Theorem 4. $\qquad \square$

## VI. TIGHTNESS OF THE BOUND OF THEOREM 3

### A. Class of programs for which the bound is tight

We provide a sufficient condition on the problems $P_k$ so that the solution returned by the scheme of Figure 2 achieves the upper bound given by the right-hand side of (23) when all the intermediate problems $P_k, k = 0, \ldots, \ell$, are fully-supported. The result of this section implies that the bound of Theorem 3 is tight, i.e., there exists a class of convex scenario programs where it holds with equality.

To this end, we replace the mapping $\mathcal{A}$ in (25) with $\bar{\mathcal{A}} : \Delta^m \to 2^\Delta$ defined

$$\bar{\mathcal{A}}(C) = \Big\{ \delta \in \Delta : g(x_\ell^\star(C), \delta) \leq 0 \Big\} \cup$$

$$\Big\{ \bigcup_{k=0}^{\ell-1} \text{supp}(x_k^\star(C)) \Big\}. \tag{55}$$

Note that $\bar{\mathcal{A}}(C)$ coincides with the one in (38), but without the set $\mathcal{A}_2(C)$ in its definition. We impose the following assumption.

**Assumption 4.** *Fix any* $S = \{\delta_1, \ldots, \delta_m\} \in \Delta^m$ *and let* $C \subset S$. *For any* $k \in \{0, \ldots, \ell\}$ *and* $\delta \in S$ *such that* $\delta \in \text{supp}(x_k^\star(C))$, *we have that*

$$g(z^\star(J), \delta) > 0,$$

*for all* $J \subset C \setminus \big( \cup_{j=0}^{k-1} \text{supp}(x_j^\star(C)) \cup \{\delta\} \big)$ *with* $|J| = d$.

Assumption 4 imposes a restriction on the class of fully-supported problems. For instance, the pictorial example of Figure 3 does not satisfy Assumption 4, even though all the intermediate problems $P_k$ are fully-supported, as the dashed-blue removed constraint is not violated by the resulting solution. Indeed, Assumption 4 requires that, with $\mathbb{P}^m$-probability

one, whenever a sample belongs to the support scenarios of any intermediate problem, then the scenario associated with it is violated by all the solutions that could have been obtained using any subset of cardinality $d$ from the remaining samples. Note that Assumption 4 is similar to the requirement of Theorem 1 [25], [26], however, in Theorem 5 below we exploit it in conjunction with the discarding scheme of Figure 2 to show that the result of Theorem 3 is tight. This serves as a constructive argument for the existential result of [25]. In this paper we do not offer any means to check the validity of Assumption 4; however, observe that the scenario programs studied in Section III satisfy such an assumption (see Remark 3 below).

**Theorem 5.** *Consider Assumptions 1, 2, and 4. Fix $\epsilon \in (0,1)$, set $r = \ell d$ and let $m > r + d$. Consider also the scenario discarding scheme as encoded by* (18) *and illustrated in Figure 2, and let the minimizer of the $\ell$-th program be $x^\star(S) = x_\ell^\star(S)$. We then have that*

$$\mathbb{P}^m \Big\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\big\{ \delta \in \Delta : g(x^\star(S), \delta) > 0 \big\} > \epsilon \Big\}$$
$$= \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \qquad (56)$$

*Proof. Existence:* We first show that the set $C$ given in (33) is a compression for the mapping in (55). Recall that under Assumption 2 we have that $R_k(S) = \text{supp}(x_k^\star(S))$ for all $k \in \{0, \dots, \ell\}$. Applying a similar induction argument as in the existence part of Theorem 3, we have that $x_k^\star(C) = x_k^\star(S)$ for all $k \in \{0, \dots, \ell\}$. Hence, by the definition of the mapping $\bar{\mathcal{A}}(C)$ in (55), we obtain that $\bar{\mathcal{A}}(C) = \bar{\mathcal{A}}(S)$, thus showing that $C$ in (33) is a compression.

*Uniqueness:* Let $C'$ be another compression of size $(\ell+1)d$. We will show that $x_k^\star(C') = x_k^\star(S)$ for all $k \in \{0, \dots, \ell\}$, which by the existence part yields that $x_k^\star(C) = x^\star(C')$ for all $k \in \{0, \dots, \ell\}$. By Assumption 1 and 2, this would then imply that $C = C'$.

To show that $x_k^\star(C') = x_k^\star(S)$ for all $k \in \{0, \dots, m\}$, it suffices to show that for all $\delta \in S \setminus C'$ we have that

$$x_k^\star(C') = x_k^\star(C' \cup \{\delta\}), \text{ for all } k \in \{0, \dots, \ell\}. \qquad (57)$$

In fact, if (57) holds for all $\delta \in S \setminus C'$ by induction it follows then that $x_k^\star(C') = x_k^\star(S)$ for all $k \in \{0, \dots, \ell\}$.

To show (57) assume for the sake of contradiction that there exist a $\bar{\delta} \in S \setminus C'$ and a $k \in \{0, \dots, \ell\}$ such that $x_k^\star(C) \neq x_k^\star(C' \cup \{\bar{\delta}\})$. Let $\bar{k}$ be the smallest index such that this occurs and note that

$$x_{\bar{k}}^\star(C') = z^\star(C' \setminus \cup_{j=0}^{\bar{k}-1} \text{supp}(x_j^\star(C'))), \qquad (58)$$

$$x_{\bar{k}}^\star(C' \cup \{\bar{\delta}\}) = z^\star((C' \setminus \cup_{j=0}^{\bar{k}-1} \text{supp}(x_j^\star(C'))) \cup \{\bar{\delta}\}), \quad (59)$$

which implies that $\bar{\delta} \in \text{supp}(x_{\bar{k}}^\star(C' \cup \{\bar{\delta}\}))$, as removal of $\bar{\delta}$ will change $x_{\bar{k}}^\star(C' \cup \{\bar{\delta}\})$ to $x_{\bar{k}}^\star(C')$. By Assumption 4 and since $\text{supp}(x_j^\star(C')) = \text{supp}(x_j^\star(C' \cup \{\bar{\delta}\}))$ for all

$j = 0, \dots, \bar{k} - 1$, we have that for all $J \subset C' \setminus \big( \cup_{j=0}^{\bar{k}-1} \text{supp}(x_j^\star(C')) \cup \{\bar{\delta}\} \big)$ with cardinality $d$,

$$g(z(J), \bar{\delta}) > 0. \qquad (60)$$

Hence, since $\bar{J} = \text{supp}(x_\ell^\star(C'))$ is a subset of cardinality $d$ of $C' \setminus \big( \cup_{j=0}^{\bar{k}-1} \text{supp}(x_j^\star(C')) \cup \{\bar{\delta}\} \big)$, as these constraints have not been removed from $C'$, we obtain that

$$g(z(\bar{J}), \bar{\delta}) = g(x_\ell^\star(C'), \bar{\delta}) > 0, \qquad (61)$$

where the equality follows from (24). However, $C'$ is assumed to be a compression set for $\bar{\mathcal{A}}$, which implies that $\delta \in \bar{\mathcal{A}}(C')$, i.e., $g(x_\ell^\star(C'), \bar{\delta}) \leq 0$. This is in contradiction with (61), implying that $x_k^\star(C') = x_k^\star(C' \cup \delta)$, for any $k \in \{0, \dots, \ell\}$, for any $\delta \in S \setminus C'$. Using induction, adding one by one $\delta \in S \setminus C'$, we can then show that $x_k^\star(C') = x_k^\star(S) = x_k^\star(C)$ for all $k \in \{0, \dots, \ell\}$, thus showing that $C$ in (33) is the unique compression set for the mapping defined in (55).

By Theorem 2, we then have that

$$\mathbb{P}^m \{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{ \delta : \delta \notin \bar{\mathcal{A}}(C) \} > \epsilon \}$$
$$= \mathbb{P}^m \{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{ \delta : g(x_\ell^\star(C), \delta) > 0 \} > \epsilon \}$$
$$= \mathbb{P}^m \{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{ \delta : g(x_\ell^\star(S), \delta) > 0 \} > \epsilon \}$$
$$= \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}, \qquad (62)$$

where the first equality follows since the union of support scenarios is a discrete set and will be of measure zero, since the problems are assumed to be fully-supported, and hence non-degenerate. To obtain the second equality we have used the fact that $x_\ell^\star(C) = x_\ell^\star(S)$ for the compression set defined in (33). This concludes the proof of Theorem 5. $\qquad \square$

**Remark 3.** *The scenario programs studied in Section III satisfy Assumption 4. This is important as it not only shows that* (6) *and* (17) *are coherent with Theorem 5 but also provides concrete examples that satisfy Assumption 4.*

## VII. NUMERICAL EXAMPLE

In this section, we consider a resource allocation problem to illustrate our theoretical result. Suppose that a manufacturer produces a good in $d$ different locations, and that this good can be produced from $n$ different resources. The quantity of resource $p$, $p = 1, \dots, n$, that is needed to produce a unitary amount of the given good at facility $j$, $j = 1, \dots, d$, is a random variable parametrized by $\delta \in \mathbb{R}$, and is denoted by $a_{pj}(\delta)$. We assume that the amount of resources $p$ available to all facilities is deterministic. The objective is to maximize the production, given by $\sum_{j=1}^{d} x^j$, where $x^j$ is the $j$-th component of $x \in \mathbb{R}^d$, while keeping the risk of running out of resources under control.

Under the scenario theory we do not have access to the distribution that generates $a_{pj}(\delta), p = 1, \dots, n, j = 1, \dots, d$; however, we encode it by means of data $(a_{pj}(\delta_i))_{i=1}^{m}$ for all
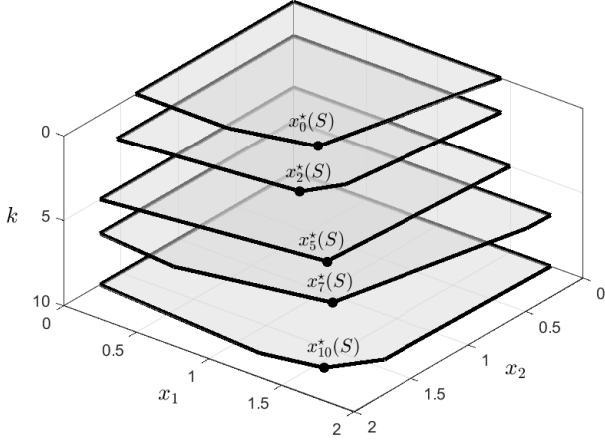
Fig. 5: Feasibility sets of the intermediate problems $P_k$, $k = 0, 2, 5, 7, 10$, for the scheme proposed in Figure 2 when applied to (63). The optimal solution of each problem is denoted by $x_k^\star(S)$, $k = 0, 2, 5, 7, 10$.
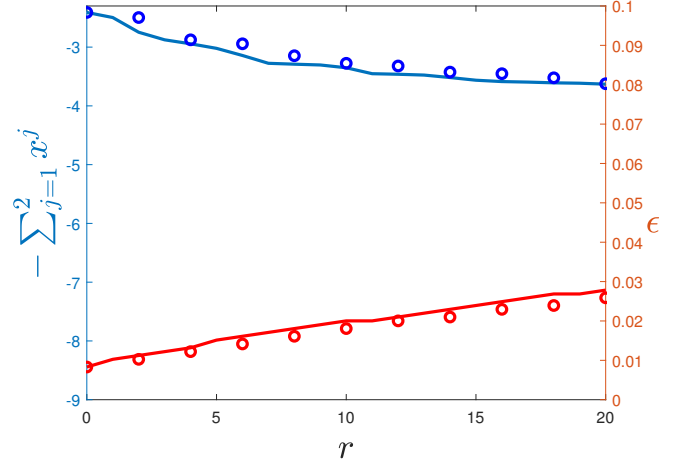


Fig. 6: Cost and probability of constraint violation for the solution returned by the scheme of Figure 2 and a greedy removal strategy for the problem in (63) when $d = 2$, $n = 2$, and $m = 2000$. With the blue dots we show the cost obtained by the proposed procedure, where we are allowed to remove scenarios in batches of $d = 2$, while the solid one shows the performance obtained by the greedy removal strategy where scenarios are removed one by one. In red we show the corresponding behavior of the probability of constraint violation. This is calculated from the bounds of Theorem 4 and Theorem 1, respectively, with $\beta = 1 \times 10^{-6}$.

$p = 1, \ldots, n$ and for all $j = 1, \ldots, d$, and solve the following fully-supported convex scenario problem

$$\begin{aligned} & \underset{\{x^j \geq 0\}_{j=1}^d}{\text{minimize}} && c^\top x \\ & \text{subject to} && A(\delta_i)x \leq b, \quad \text{for all } i = 1, \ldots, m, \end{aligned} \quad (63)$$

where, for each $i \in \{1, \ldots, m\}$, $A(\delta_i) \in \mathbb{R}^{n \times d}$ is a matrix whose $(p, j)$-th entry is given by $a_{pj}(\delta_i)$, $b \in \mathbb{R}^n$ is a vector whose $p-$th component is the amount of resource $p$ available to all facilities, and $c = \begin{bmatrix} -1 & \cdots & -1 \end{bmatrix}^\top \in \mathbb{R}^d$.

### A. The two-dimensional case

Set $d = 2$ and consider 2000 scenarios from the unknown distribution[1] for $\delta$. We study the behavior of the scheme in Figure 2 when we discard $r = 20$ of these scenarios. In this case, note that according to the description given in Section IV, we have to solve a cascade of 11 optimization problems (i.e, $\ell = 10$ in the scheme of Figure 2).

Figure 5 illustrates the feasible set for stages $k = 0, 2, 5, 7$, and 10 of the scheme of Figure 2, and depicts the corresponding optimal solution for each $P_k$ as $x_k^\star(S)$. Note that the feasible set associated to each problem $P_k$ grows as we remove scenarios. To complement this analysis, we also show in Figure 6 a comparison between our method and the greedy scenario removal strategy as described in [25], which removes scenarios one by one according to one that yields the best improvement in the cost. With the blue dots we show the cost obtained by the proposed procedure, where we are allowed to remove scenarios in batches of $d = 2$, while the solid one shows the performance obtained by the greedy removal

strategy, where scenarios are removed one by one. In red we show the corresponding behavior of the probability of constraint violation. This is calculated from the bounds of Theorem 3 and Theorem 1, respectively, with $\beta = 1 \times 10^{-6}$.

### B. The 10-dimensional case

Consider now (63) with $d = 10$ and the same 2000 scenarios. We compare the cost improvement of the proposed bound (Theorem 4) with the one of Theorem 1 [25]. To this end, for a given $\epsilon \in [0.01, 0.08]$, we compute the maximum number of scenarios that can be removed using each of these bounds. Note that due to the fact that we remove scenarios is batches of $d$, we compute the number of scenarios that need to be removed by means of numerical inversion from the bound of Theorem 4 (using $m = 2000$, $\beta = 1 \times 10^{-6}$ and the given $\epsilon$), and round it down to the closest multiple of $d = 10$. For instance, for $\epsilon = 0.03$ the maximum number of scenarios that can be removed using the bound in (4) is $r = 18$, but we only remove 10. Figure 7 shows then the relative cost difference $100 \times \frac{f^\star(\epsilon) - \bar{f}^\star(\epsilon)}{\bar{f}^\star(\epsilon)}$ as a function of $\epsilon$, where $f^\star(\epsilon)$ is the optimal value of problem (63) when scenarios are removed according to Theorem 4, and $\bar{f}^\star(\epsilon)$ correspond to the bound in [25]. Note that for $\epsilon > 0.03$ our scheme results in an improvement in performance (i.e., the relative cost difference is negative) achieving approximately $4\%$ improvement when $\epsilon = 0.08$. This is due to the fact that more scenarios can be removed, while guaranteeing the same level of violation. Notice also that there is no improvement when $\epsilon \leq 0.03$. This

---

[1] For our simulations, fix $i \in \{1, \ldots, m\}$ and generate an auxiliary matrix, $B(\delta_i) \in \mathbb{R}^{n \times d}$, whose entries are obtained from a Laplacian distribution with mean equal to one and variance equal to three. Then set $A(\delta_i) = 0.04B(\delta_i)$. Our numerical results were obtained setting the "seed" equal to 30 in MATLAB.
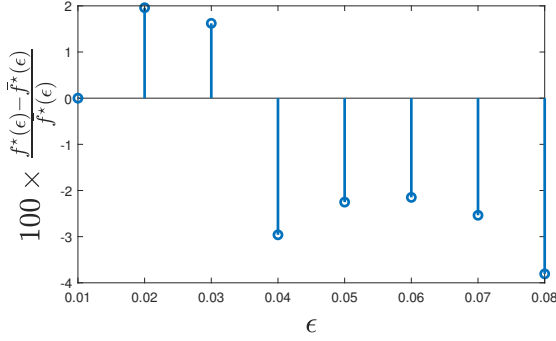
Fig. 7: Relative cost improvement $100 \times \frac{f^\star(\epsilon) - \bar{f}^\star(\epsilon)}{\bar{f}^\star(\epsilon)}$, as a function of $\epsilon$, where $f^\star(\epsilon)$ corresponds to the cost associated with Theorem 4, and $\bar{f}^\star(\epsilon)$ to the one of Theorem 1, [25]. The numerical results correspond to (63) with $d = 10$.

is due to the limitation on the number of removed scenarios, as for $\epsilon \in [0.01, 0.03]$ the proposed bound returns a value for $r$ that is less than 10, hence no scenarios are removed.

The computational requirements of the proposed approach are lower with respect to the greedy removal strategy in [25] (see also [26]). To put this in perspective, to remove 100 scenarios in the previous example when $d = 10$, the greedy strategy requires the solution of 1101 optimization problems of the form (63), whereas the proposed scheme only needs to solve 11 of these problems. The computational savings are more pronounced as the dimension of the problem grows.

**Remark 4.** *Theorem 4 offers an improvement with respect to Theorem 1, [25], [26], as far as guarantees on the probability of constraint violation are concerned. However, the effect of the proposed discarding scheme with respect to the greedy removal strategy described in [25], [26], does not necessarily follow the one of Figure 7, but is problem dependent.*

## VIII. CONCLUDING REMARKS

In this paper we revisited the sampling and discarding approach for scenario based optimization. We proposed a scenario discarding scheme that consists of a cascade of optimization problems, where at each stage we remove a superset of the support constraints. By relying on results from compression learning theory, we provide a less conservative bound on the probability of constraint violation of the obtained solution. Besides, we show that the proposed bound is tight, and characterize a class of problems for which this is the case. Our results are illustrated on a resource sharing linear program, as well as by means of two examples that admit analytic solutions.

Current work concentrates towards extending our scenario discarding scheme so that we no longer remove scenarios in batches but one by one. Preliminary results in this direction can be found in [45]. Moreover, we aim at exploiting the dual variables associated with each constraint in order to create a tie-break rule to choose the scenarios to be removed at each stage.

## REFERENCES

[1] M. Vidyasagar, "Statistical Learning Theory and Randomized Algorithms for Control," *IEEE Control Systems Magazine*, vol. 18, no. 6, pp. 69–85, 1998.

[2] T. Alamo, R. Tempo, and E. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2545–2559, 2009.

[3] K. Margellos, P. Goulart, and J. Lygeros, "On the road between robust optimization and the scenario approach for chance constrained optimization problems," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2258–2263, 2014.

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

[5] G. Calafiore, "Repetitive Scenario Design," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1125–1137, 2017.

[6] A. Ramponi, "Consistency of the scenario approach," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 135–162, 2018.

[7] M. Campi, S. Garatti, and F. Ramponi, "A General Scenario Theory for Nonconvex Optimization and Decision Making," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4067–4078, 2018.

[8] M. Campi and S. Garatti, "Wait-and-judge scenario optimization," *Mathematical Programming*, vol. 167, no. 1, pp. 155–189, 2018.

[9] D. Bertsimas and M. Sim, "Tractable approximations to robust conic optimization problems," *Mathematical Programming*, vol. 107, no. 1-2, pp. 5–36, 2006.

[10] D. Bertsimas, D. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011.

[11] A. Ben-Tal and A. Nemirovski, "Selected topics in robust convex optimization," in *Mathematical Programming*, vol. 112, no. 1, 2008, pp. 125–158.

[12] A. Ben-Tal, L. Ghaoui, and A. Nemirovski, *Robust optimization*, 1st ed. Princeton University Press, 2009.

[13] M. de Oliveira, J. Bernussou, and J. Geromel, "A new discrete-time robust stability condition," *Systems and Control Letters*, vol. 37, no. 4, pp. 261–265, jul 1999.

[14] Y. Fujisaki, F. Dabbene, and R. Tempo, "Probabilistic design of LPV control systems," *Automatica*, vol. 39, no. 8, pp. 1323–1337, 2003.

[15] J. Geromel, M. de Oliveira, and J. Bernussou, "Robust filtering of discrete-time linear systems with parameter dependent Lyapunov functions," *SIAM Journal on Control and Optimization*, vol. 41, no. 3, pp. 700–711, 2003.

[16] G. Calafiore, F. Dabbene, and R. Tempo, "Research on probabilistic methods for control system design," *Automatica*, vol. 47, no. 7, pp. 1279–1293, 2011.

[17] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*, 2nd ed. Springer, London, 2013.

[18] L. Romao, M. de Oliveira, P. Peres, and R. Oliveira, "State-feedback and filtering problems using the generalized KYP lemma," in *Proceedings of the IEEE International Symposium on Computer-Aided Control System Design*, oct 2016, pp. 1054–1059.

[19] A. Prékopa, "On Probabilistic Constrained Programming," in *Proceedings of the Princeton Symposium on Mathematical Programming*, 2015, pp. 113–138.

[20] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 959–996, 2006.

[21] B. Pagnoncelli, S. Ahmed, and A. Shapiro, "Sample average approximation method for chance constrained programming: Theory and applications," *Journal of Optimization Theory and Applications*, vol. 142, no. 2, pp. 399–416, 2009.

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

[23] ——, "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742—-753, 2006.

[24] M. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1211–1230, 2008.

[25] ——, "A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality," *Journal of Optimization Theory and Applications*, vol. 148, pp. 257–280, 2011.

[26] G. Calafiore, "Random convex programs," *SIAM Journal on Optimization*, vol. 20, no. 6, pp. 3427–3464, 2010.

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

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

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

[30] G. Calafiore and L. Fagiano, "Robust model predictive control via scenario optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 219–224, 2013.

[31] Y. Ma, J. Matusko, and F. Borrelli, "Stochastic model predictive control for building HVAC systems: Complexity and conservatism," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 1, pp. 101–116, jan 2015.

[32] B. Varga, T. Tettamanti, B. Kulcsár, and X. Qu, "Public transport trajectory planning with probabilistic guarantees," *Transportation Research Part B: Methodological*, vol. 139, pp. 81–101, sep 2020.

[33] X. Shen, T. Ouyang, Y. Zhang, and X. Zhang, "Computing Probabilistic Bounds on State Trajectories for Uncertain Systems," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–6, sep 2020.

[34] J. Kudela and P. Popela, "Pool Discard Algorithm for Chance Constrained Optimization Problems," *IEEE Access*, vol. 8, pp. 79 397–79 407, 2020.

[35] A. Devonport and M. Arcak, "Estimating Reachable Sets with Scenario Optimization," *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, no. 120, pp. 75–84, jul 2020.

[36] M. Campi and S. Garatti, "Scenario optimization with relaxation: a new tool for design and application to machine learning problems," in *59th Conference on Decision and Control*, 2020, pp. 2463—-2468.

[37] M. Cubuktepe, N. Jansen, S. Junges, J. Katoen, and U. Topcu, "Scenario-based verification of uncertain MDPS," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12078 LNCS. Springer, apr 2020, pp. 287–305.

[38] K. Jensen and A. Podelski, "Tools and algorithms for the construction and analysis of systems," Cham, pp. 177–179, 2006.

[39] M. Shahroz, M. Younis, and H. Nasir, "A Scenario-Based Stochastic Optimization Approach for Non-Intrusive Appliance Load Monitoring," *IEEE Access*, vol. 8, pp. 142 205–142 217, 2020.

[40] A. Scampicchio, A. Aravkin, and G. Pillonetto, "Stable and Robust LQR Design via Scenario Approach," 2020.

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

[42] S. Floyd and M. Warmuth, "Sample compression, learnability, and the Vapnik-Chervonenkis dimension," *Machine Learning*, vol. 21, no. 21, pp. 269–304, 1995.

[43] D. Salamon, *Measure and Integration*. Zurich, Switzerland: European Mathematical Society, 2016, 2016.

[44] L. Romao, K. Margellos, and A. Papachristodoulou, "Tight generalization guarantees for the sampling and discarding approach to scenario optimization," in *59th IEEE Conference on Decision and Control*, 2020, pp. 2228–2233.

[45] ——, "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. To appear. [Online]. Available: http://sysos.eng.ox.ac.uk/wiki/images/4/46/RMP20eL4DC.pdf

**Licio Romao** received the B.S. degree from the Federal University of Campina Grande - UFCG, Brazil, in 2014, and M.S. degree from the University of Campinas - UNICAMP, Brazil, in 2017. He is currently pursuing the Ph.D degree in the Control Group, Department of Engineering Science, University of Oxford, Oxford, United Kingdom.

He visited the University of Rome Tor Vergata, Italy, in 2012, the University of California San Diego, CA, USA, in 2015, and the University of Bologna, Italy, in 2019. His research interests include optimization algorithms and control strategies applied to large-scale, uncertain systems.

**Antonis Papachristodoulou** FIEEE received the M.A./M.Eng. degree in electrical and information sciences from the University of Cambridge, Cambridge, U.K., and the Ph.D. degree in control and dynamical systems (with a minor in aeronautics) from the California Institute of Technology, Pasadena, CA, USA. He is currently Professor of Engineering Science at the University of Oxford, Oxford, U.K., and a Tutorial Fellow at Worcester College, Oxford. He is also an EPSRC Fellow for Growth in Synthetic Biology and the Director of the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology. His research interests include large-scale nonlinear systems analysis, sum of squares programming, synthetic and systems biology, networked systems, and flow control. Professor Papachristodoulou received the 2015 European Control Award for his contributions to robustness analysis and applications to networked control systems and systems biology. In the same year, he received the O. Hugo Schuck Best Paper Award.

**Kostas Margellos** received the Diploma in electrical engineering from the University of Patras, Patras, Greece, in 2008, and the Ph.D. degree in control engineering from ETH Zurich, Zurich, Switzerland, in 2012.

He spent 2013, 2014 and 2015 as a Postdoctoral Researcher at ETH Zurich, UC Berkeley, and Politecnico di Milano, respectively. In 2016, he joined the Control Group, Department of Engineering Science, University of Oxford, Oxford, U.K., where he is currently an Associate Professor. He is also a Lecturer at Worcester College, Oxford. His research interests include optimization and control of complex uncertain systems, with applications to generation and load side control for power networks.