

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

Licio Romao, Kostas Margellos, and Antonis Papachristodoulou

**Abstract**—We revisit the so-called sampling and discarding approach used to quantify the probability of violation of a scenario solution when some of the original samples are allowed to be discarded. We propose a 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 produce a tighter bound for the probability of violation of the obtained solution than existing state-of-the-art one. Besides, we show that the proposed bound is tight by exhibiting a class of optimization problems that achieves the given upper bound. The improvement of the proposed methodology with respect to a scenario discarding scheme based on a greedy removal strategy is shown by means of an analytic example and 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 problem 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.

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

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

One of the fundamental development 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.

To alleviate conservatism, 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. As opposed to the original bound in [24], however, the bound on the probability of constraint violation in [25], [26] is not tight. In particular, in [25], the authors provide an existential statement, showing that there exists a problem class and a scenario removal algorithm for which a tighter bound may be achieved. However, the corresponding analysis is not constructive.

In this paper, we revisit the sampling and discarding methodology and provide a constructive way to achieve the tighter bound discussed in [25], thus improving upon the main theorems in [25], [26]. Our approach consists of solving a cascade of optimization problems, and removing a subset of certain cardinality from the original samples at each stage of the process. Similarly to [25], [26], our bound on the probability of constraint violation, under a non-degeneracy assumption, holds for all convex problems.

In particular, our contributions can be summarized as follows: (1) we propose a specific removal scheme and show that it possesses tighter guarantees on the probability of constraint violation associated to the resulting solution compared to [25], [26]; (2) we relax the assumption present in the papers [25], [26] that the removed scenarios need to be violated by the resulting solution; (3) we provide a problem class that satisfies our bound with equality, thus proving that the obtained bound is tight for this particular removal scheme. We also illustrate this statement by an example that admits an analytic solution; (4) our proof line is novel, in the sense that it departs from the one of [25], and is based on probably approximately correct learning (PAC) learning concepts based on the notion of compression [27], [30], [31].

Note that our results are based on a particular scenario discarding scheme, that requires removing scenarios in batches with cardinality equal to the dimension of the problem, 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], [32], 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 introduces the proposed scenario discarding scheme and states the main results of the paper, while their proofs are provided in Section IV. Section V provides a class of optimization programs for which the proposed result is tight, while Section VI illustrates the theoretical results by means of a numerical example. Finally Section VII 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  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  is a probability measure on  $\Delta$  (see [33] for more details). Fix any  $m \in \mathbb{N}$ , and let  $S = \{\delta_1, \delta_2, \dots, \delta_m\}$  be independent and identically distributed (i.i.d.) samples from  $\mathbb{P}$ . Note that  $(\delta_1, \dots, \delta_m) \in \Delta^m$ ; 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  $\Delta$  and  $\mathbb{P}$  are 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 given a collection of  $S$  scenarios, constraints are enforced only on those scenarios [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] (it is also referred to as scenario approach with constraint removal [26]).

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

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

where  $x \in \mathbb{R}^d$ ,  $X$  is a closed and convex set of  $\mathbb{R}^d$ , and the function  $g : \mathbb{R}^d \times \Delta \rightarrow \mathbb{R}$  is convex in  $x$  for all  $\delta \in \Delta$ . Note that the scenarios in the set  $R$  have been removed. 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  $x^*(S)$  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, even though in (1),  $R$  of them are removed. The following result from [25] characterizes the probability that  $x^*(S)$  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  $x^*(S)$  the optimal solution of (1). If all removed scenarios are violated by the resulting solution  $x^*(S)$ , i.e.,  $g(x^*(S), \delta) > 0$  for all  $\delta \in R$ , with  $\mathbb{P}^m$ -probability one, then*

$$\begin{aligned} & \mathbb{P}^m \left\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P} \{ \delta \in \Delta : g(x^*(S), \delta) > 0 \} > \epsilon \right\} \\ & \leq \binom{r+d-1}{r} \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \end{aligned} \quad (2)$$

Note 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  $x^*(S)$ . This is instrumental in the proof of Theorem 1, as shown in [25], [26]. Naturally, if scenarios were discarded in an optimal way, this may lead to a better improvement in performance compared to any other removal scheme. However, the optimal discarding scheme is of combinatorial complexity, and typically one resorts to a greedy removal algorithm [25], [26].

Moreover, note that the bound of Theorem 1 is not tight; 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 both as far as the discarding scheme and the problem class is concerned.

We will propose in Section III an alternative discarding strategy that possesses tighter guarantees for its final solution than the one of Theorem 1, at the expense of restricting the allowable number of discarded scenarios. Our scheme also relaxes the assumption of Theorem 1 that requires all the removed scenarios to be violated by the final solution.

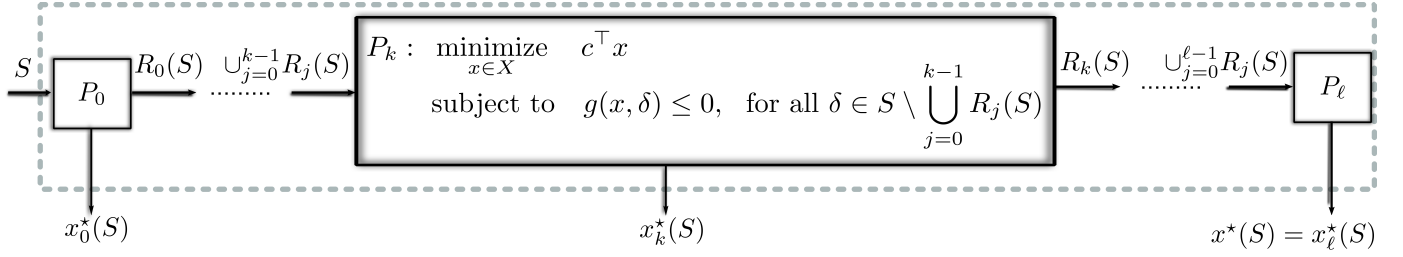


Fig. 1: Block diagram of the proposed scheme. For a given set of scenarios  $S = \{\delta_1, \dots, \delta_m\}$  with  $(\ell + 1)d < m$ , we solve a cascade of  $\ell + 1$  optimization programs denoted by  $P_k$ ,  $k \in \{0, \dots, \ell\}$ . For each  $k$ ,  $k \in \{0, \dots, \ell - 1\}$ , we remove  $R_k(S)$  scenarios with  $|R_k(S)| = d$ , hence, in total  $r = \ell d$  scenarios (the ones in  $\bigcup_{j=0}^{\ell-1} R_j(S)$ ) are discarded. The choice of each set of discarded scenarios depends on the initial set  $S$ , thus we introduce it as argument of  $R_k$ . If each problem is fully-supported, the elements of  $R_k(S)$  correspond to the (unique) set of support scenarios associated with the minimizer  $x_k^*(S)$  of that program – see (4); otherwise,  $R_k(S)$  contains the support scenarios as well as additional scenarios selected according to a lexicographic order regularization procedure, as in (7). The final solution is denoted by  $x^*(S) = x_\ell^*(S)$ .

### B. Learning theoretic concepts

In this section we provide some background learning theoretic results on which we leverage for proving the main results of this paper. This relies on the following definition.

**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  $\mathcal{A} : \Delta^m \rightarrow 2^\Delta$ . If*

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

*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 also known as consistency of  $\mathcal{A}(C)$  with respect to the samples [27], [30]. The main focus within a probably approximately correct (PAC) learning framework (see [30] and references therein) is to study the generalization properties of  $\mathcal{A}(C)$ , i.e., 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*

$$\mathbb{P}^m \left\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P} \{ \delta \in \Delta : \delta \notin \mathcal{A}(C) \} > \epsilon \right\} = \sum_{i=0}^{\zeta-1} \binom{m}{i} \epsilon^i (1 - \epsilon)^{m-i}. \quad (3)$$

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 this property is called PAC within the learning literature. Hence, we can reinterpret the result of Theorem 2 as stating 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. PROPOSED DISCARDING SCHEME AND MAIN RESULTS

In this section, we present a particular scenario discarding scheme which does not require that all removed scenarios are violated by the resulting solution. For a given set of scenarios  $S = \{\delta_1, \dots, \delta_m\}$ , we solve a cascade of  $\ell + 1$ , where  $(\ell + 1)d < m$ , optimization programs denoted by  $P_k$ ,  $k \in \{0, \dots, \ell\}$ . For each  $k \in \{1, \dots, \ell\}$ , the corresponding program is given by

$$P_k : \begin{aligned} &\text{minimize}_{x \in X} && c^\top x \\ &\text{subject to} && g(x, \delta) \leq 0, \text{ for all } \delta \in S \setminus \bigcup_{j=0}^{k-1} R_j(S), \end{aligned}$$

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  $\bigcup_{j=0}^{\ell-1} R_j(S)$ ). 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 1.

Our choice for  $R_k(S)$ ,  $k \in \{0, \dots, \ell - 1\}$ , will be detailed in the following two subsections, and it relies on the properties of each minimizer  $x_k^*(S)$ .

**Definition 2** (Support set; see Definition 2 in [24]). *Fix any  $k \in \{0, \dots, \ell\}$  and consider  $P_k$ . An element of  $S \setminus \bigcup_{j=0}^{k-1} R_j(S)$  is said to be a support scenario of  $P_k$ , if its removal changes*

the minimizer  $x_k^*(S)$ . The support set of  $x_k^*(S)$ , denoted by  $\text{supp}(x_k^*(S))$ , is the collection of support scenarios of  $S \setminus \bigcup_{j=0}^{k-1} R_j(S)$ .

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

**Definition 4** (Non-degenerate programs; see Assumption 2 in [8]). Fix any  $k \in \{0, \dots, \ell\}$  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,  $\text{supp}(x_k^*(S))$ , results in  $x_k^*(S)$ , 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 the dimension of the decision vector  $d$  [24], [27].

We distinguish two cases according to whether the underlying problem is fully-supported or only non-degenerate.

#### A. The fully-supported case

Throughout this section the cardinality of the support set of problem  $P_k$ ,  $k = 0, \dots, \ell$ , is assumed to be equal to  $d$ , which is the dimension of the optimization variable, with  $\mathbb{P}^m$ -probability one. We formalize this statement in the following assumption.

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

Under Assumption 2, our choice for the removed scenarios is given by

$$R_k(S) = \text{supp}(x_k^*(S)), \quad k \in \{0, \dots, \ell - 1\}, \quad (4)$$

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  $x_\ell^*(S)$ ; this quantity will be used in the sequel. Note that  $R_\ell(S)$  does not contain any removed scenarios.

One of the main results of this paper is to tighten the bound of Theorem 1 under Assumption 2. This is achieved 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 (4) and illustrated in Figure 1,

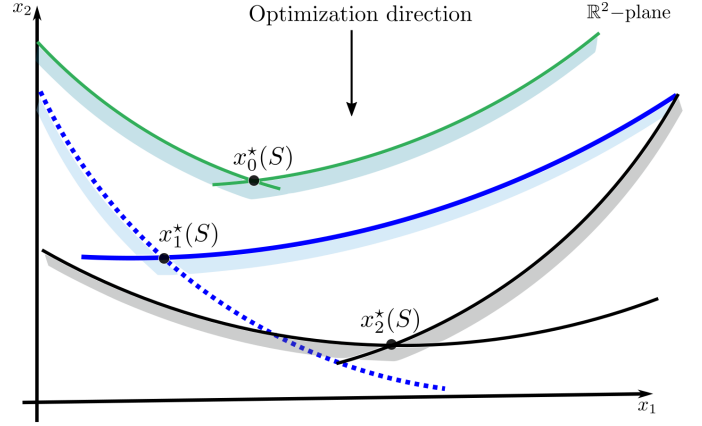


Fig. 2: Pictorial example that illustrates the scheme proposed in Figure 1 for fully-supported problems. In this case, we have that  $d = 2$ ,  $m = 6$ ,  $r = 4$ , and  $\ell = 2$ , and all the problems  $P_k$ ,  $k \in \{0, 1, 2\}$ , satisfy Assumption 2. The objective function is given by  $c^\top x = x_2$  (indicated by the downwards pointing arrow). The constraint sets are denoted by the different colors patterns: the green constraints are associated to the samples of  $\text{supp}(x_0^*(S))$ , the blue ones to  $\text{supp}(x_1^*(S))$ , and the black ones correspond to  $\text{supp}(x_2^*(S))$ . Observe that the dashed-blue constraint is removed by the scheme of Figure 1, but it is not violated by  $x_2^*(S)$ .

and let the minimizer of the  $\ell$ -th program be  $x^*(S) = x_\ell^*(S)$ . We then have that

$$\mathbb{P}^m \left\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^*(S), \delta) > 0\} > \epsilon \right\} \leq \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1 - \epsilon)^{m-i}. \quad (5)$$

The proof of Theorem 3 is deferred to Section IV-A. Note that unlike [25], [26], we do not require the removed scenarios to be violated by the resulting solution (see Figure 2).

To illustrate how the proposed scheme works, we consider the pictorial example of Figure 2. Note that  $d = 2$ ,  $m = 6$ , and we remove  $r = 4$ , thus requiring  $\ell = 2$  steps of the removal scheme of Figure 1. All the problems  $P_k$ ,  $k \in \{0, 1, 2\}$ , are fully-supported, thus satisfying Assumption 2. The objective function is given by  $c^\top x = x_2$  and is indicated by the downwards pointing arrow. The corresponding solution for the intermediate problems is illustrated by  $x_k^*(S)$ , 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,  $\text{supp}(x_0^*(S))$ , 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  $x_1^*(S)$  violates both constraints that belong to  $\text{supp}(x_0^*(S))$  and satisfies all the remaining ones. The result of Theorem 3 provides guarantees for the probability of violation of  $x_2^*(S)$ . 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  $P_k$ ,  $k \in \{0, \dots, \ell\}$ , is non-degenerate.

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

In case of a non fully-supported problem ( $\text{supp}(x_k^*(S)) < d$ , for some  $k \in \{0, \dots, \ell\}$ ), 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  $\text{supp}(x_k^*(S))$ , 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  $\text{supp}(x_k^*(S))$  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  $k \in \{0, \dots, \ell\}$ , let  $\nu_k(S) = d - |\text{supp}(x_k^*(S))|$  and define the following sets recursively as

$$Z_k(S) = \left\{ \nu_k(S) \text{ scenarios with the smallest labels in } S \setminus \left( \bigcup_{j=0}^{k-1} \{ \text{supp}(x_j^*(S)) \cup Z_j(S) \} \cup \text{supp}(x_k^*(S)) \right) \right\}, \quad (6)$$

with  $Z_0(S)$  containing the  $\nu_0(S)$  smallest according to the linear order elements of  $S \setminus \text{supp}(x_0^*(S))$ . Note that the set appearing in the definition of  $Z_k(S)$  in (6) corresponds to scenarios available at stage  $k$  that are not of support.

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

$$R_k(S) = \text{supp}(x_k^*(S)) \cup Z_k(S). \quad (7)$$

Notice that by construction  $|R_k(S)| = d$ , while if for any  $k \in \{0, \dots, \ell\}$ ,  $P_k$  is fully-supported, then  $R_k(S) = \text{supp}(x_k^*(S))$ , i.e., it coincides with the support set of  $x_k^*(S)$ . Similar as in the fully-supported case, we denote by  $R_\ell(S)$  the superset of the support set of  $x_\ell^*(S)$  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  $x_k^*(S)$ .

**Remark 1.** Consider two arbitrary scenario sets  $C \subset C'$ , and denote by  $x_k^*(C)$  and  $x_k^*(C')$  the minimizers of  $P_k$  with  $C$  and  $C'$ , respectively, replacing  $S$ . Moreover, define  $Z_k(C)$  and  $Z_k(C')$  as in (6) with  $C$  and  $C'$ , respectively, in place of  $S$ . We then have that  $F_k(x_k^*(C)) < F_k(x_k^*(C'))$  if: either  $c^\top x_k^*(C) < c^\top x_k^*(C')$ ; or  $c^\top x_k^*(C) = c^\top x_k^*(C')$  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

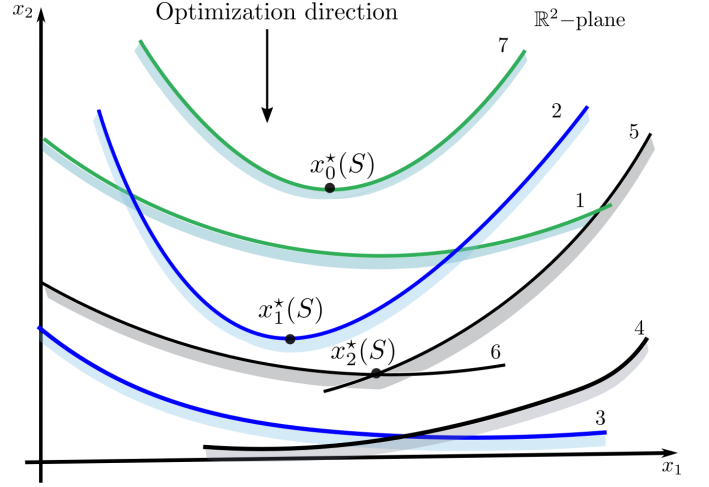


Fig. 3: Illustration of the proposed scheme applied to non-degenerate, but not fully-supported, problems. The intermediate solutions are denoted by  $x_k^*(S)$ ,  $k = 0, 1, 2$ . The different colour patterns depict the removed scenarios at each stage. The green constraints are the ones removed at stage 0, the blue ones those removed at stage 1 of the scheme presented in Figure 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.

procedure in Section IV-B to prove Theorem 4 below. It is shown in [26], that  $P_k$  with its objective function replaced by

$$F_k(x) = (c^\top x, Z_k(S)). \quad (8)$$

is a fully-supported program, and the constructed set  $R_k(S)$  in (7) 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 (7) and illustrated in Figure 1, and let the minimizer of the  $\ell$ -th program be  $x^*(S) = x_\ell^*(S)$ . We then have that

$$\mathbb{P}^m \left\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P} \{ \delta \in \Delta : g(x^*(S), \delta) > 0 \} > \epsilon \right\} \leq \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1 - \epsilon)^{m-i}. \quad (9)$$

It is important to note that (9) holds for any linear order imposed in the original samples  $S$ . Note that the optimal objective value of the scheme, however, depends on the imposed linear order, and we only provide feasibility guarantees and not optimality. However, this is also the case for the greedy strategy 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.

**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 1 works when applied to non-degenerate problems, consider the example depicted in Figure 3. Similar as before, we have  $d = 2, m = 7$ , and want to remove 4 constraints, i.e.,  $r = 4$ . As opposed to Figure 2, 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 (6). 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  $x_0^*(S)$  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 3. Then, we solve problem  $P_1$  with the resulting scenarios, obtaining  $x_1^*(S)$  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  $x_1^*(S)$ , 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  $x_2^*(S)$ .

#### IV. 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 1, fed by  $C$  rather than  $S$ . All quantities introduced in Section III depending on  $S$  would now depend on  $C$  instead. For a given set of indices  $I \subset C$ , we define

$$\begin{aligned} z^*(I) = \operatorname{argmin}_{x \in X} \quad & c^\top x \\ \text{subject to} \quad & g(x, \delta) \leq 0, \text{ for all } \delta \in I. \end{aligned} \quad (10)$$

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

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

$$\begin{aligned} \mathcal{A}(C) = & \left\{ \delta \in \Delta : g(x_\ell^*(C), \delta) \leq 0 \right\} \\ & \cap \left\{ \bigcap_{k=0}^{\ell} \left\{ \delta \in \Delta : c^\top z^*(J \cup \{\delta\}) \leq c^\top x_k^*(C), \right. \right. \\ & \quad \left. \left. \text{for all } J \subset C \setminus \bigcup_{j=0}^{k-1} R_j(C), \text{ with } |J| = d-1 \right\} \right\} \\ & \cup \left\{ \bigcup_{k=0}^{\ell-1} R_k(C) \right\} = \left( \mathcal{A}_1(C) \cap \mathcal{A}_2(C) \right) \cup \mathcal{A}_3(C). \end{aligned} \quad (11)$$

The main motivation to define the mapping in (11) is the fact that its probability of violation will be shown to upper bound that of  $\{\delta \in \Delta : g(x_\ell^*(C), \delta) \leq 0\}$ , which is ultimately the quantity we are interested in (as shown in Section IV-A, step 3).

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

- i)  $\mathcal{A}_1(C)$  contains all realizations of  $\delta$  for which the final decision of our proposed scheme  $x_\ell^*(C) = x^*(C)$  remains feasible. This is the set whose probability of occurrence we are ultimately interested to bound.
- ii)  $\mathcal{A}_2(C)$ , the intersection of  $\ell+1$  sets, indexed by  $k \in \{0, \dots, \ell\}$ , each of them containing the realizations of  $\delta$  such that, for all subsets of cardinality  $d-1$  from the remaining samples at stage  $k$ , the cost  $c^\top z^*(J \cup \{\delta\})$  is lower than or equal to  $c^\top x_k^*(C)$ . The former cost corresponds to appending  $\delta$  to any set  $J$  of  $d-1$  scenarios from  $C \setminus \bigcup_{j=0}^{k-1} R_j(C)$ , while the latter corresponds to the cost of the minimizer  $x_k^*(C)$  of  $P_k$ . Informally, this inequality is of similar nature with that of the first set in  $\mathcal{A}(C)$ , however, rather than considering constraint satisfaction it only involves some cost dominance condition for each of the interim and the final optimal solutions. The motivation to use this representation rather than constraint satisfaction conditions stems from the fact that in Section III-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)  $\mathcal{A}_3(C)$ , which includes all scenarios that are removed by the discarding scheme. Implicit in the definition of mapping (11) 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 (11).

**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 (11) if and only if, for all  $k \in \{0, \dots, \ell\}$ ,

$$x_k^*(C) = x_k^*(S). \quad (12)$$

*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

$$x_k^*(C) \neq x_k^*(C \cup \{\bar{\delta}\}). \quad (13)$$

Let  $\bar{k}$  be the minimum index such that (13) holds, while we have that  $x_j^*(C) = x_j^*(C \cup \{\bar{\delta}\})$ , for all  $j < \bar{k}$ .

By Assumption 2, the last statement implies that  $\text{supp}(x_j^*(C)) = \text{supp}(x_j^*(C \cup \{\bar{\delta}\}))$ , 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  $R_j(C) = \text{supp}(x_j^*(C))$  for fully-supported problems (similarly for  $R_j(C \cup \{\bar{\delta}\})$ ). By (10), we then have

$$x_k^*(C) = z^*(C \setminus \cup_{j=0}^{\bar{k}-1} R_j(C)), \quad (14)$$

$$x_k^*(C \cup \{\bar{\delta}\}) = z^*((C \setminus \cup_{j=0}^{\bar{k}-1} R_j(C)) \cup \{\bar{\delta}\}). \quad (15)$$

Since the right-hand side of (15) involves one more scenario with respect to the right-hand side of (14), the feasible set of (15) is a subset set of the one of (14). Moreover, by the fact that  $x_k^*(C \cup \{\bar{\delta}\}) \neq x_k^*(C)$  and Assumption 1, we obtain that

$$c^\top x_k^*(C) < c^\top x_k^*(C \cup \{\bar{\delta}\}). \quad (16)$$

Notice that  $\bar{\delta}$  belongs to the support set of  $x_k^*(C \cup \{\bar{\delta}\})$ , as its removal results in a different optimal solution with lower cost in (16). In other words, there exists  $\bar{J} \subset C \setminus \cup_{j=0}^{\bar{k}-1} R_j(C)$  (in fact,  $\bar{J} = \text{supp}(x_k^*(C \cup \{\bar{\delta}\})) \setminus \{\bar{\delta}\}$ ) of cardinality  $d-1$  such that by (15), we have that

$$c^\top x_k^*(C) < c^\top x_k^*(C \cup \{\bar{\delta}\}) = c^\top z^*(\bar{J} \cup \{\bar{\delta}\}). \quad (17)$$

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  $k \leq \ell$ ,

$$c^\top z^*(J \cup \{\bar{\delta}\}) \leq c^\top x^*(C) \leq c^\top x_k^*(C), \quad (18)$$

where the first inequality follows from the fact that  $c^\top x^*(C)$  is the optimal value for  $P_\ell$ , and  $x^*(C) = x_\ell^*(C)$  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 (18) establishes a contradiction with (17), thus showing that  $x_k^*(C) = x_k^*(C \cup \{\bar{\delta}\})$ , 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  $x_k^*(C) = x_k^*(S)$ , 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  $x_k^*(C) = x_k^*(S)$  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 (11) 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  $x_k^*(C) = x_k^*(S)$  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

$$C = \bigcup_{k=0}^{\ell} \text{supp}(x_k^*(S)), \quad (19)$$

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

*Existence:* We prove that  $C$  in (19) is a compression set. By the sufficient part of Proposition 1, it suffices to show that the set  $C$  in (19) satisfies  $x_k^*(C) = x_k^*(S)$ , for all  $k \in \{0, \dots, \ell\}$ . We will show this by means of induction. For the base case  $k = 0$ , notice that

$$\begin{aligned} c^\top x_0^*(S) &= c^\top z^*(S) = c^\top z^*(\text{supp}(x_0^*(S))) \\ &= c^\top x_0^*(C), \end{aligned} \quad (20)$$

where the first equality is due to (10), the second equality is due to the fact that  $\text{supp}(x_0^*(S))$  is the support set of  $x_0^*(S)$ , while the last equality is due to Assumption 2, the definition of support set and the fact that  $\text{supp}(x_0^*(S)) \subset C$ . By (20), and Assumption 1, we conclude that  $x_0^*(C) = x_0^*(S)$ .

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

$$\begin{aligned} c^\top x_{\bar{k}+1}^*(C) &= c^\top z^*(C \setminus \cup_{j=0}^{\bar{k}} R_j(S)) \\ &\leq c^\top z^*(S \setminus \cup_{j=0}^{\bar{k}} R_j(S)) = c^\top x_{\bar{k}+1}^*(S), \end{aligned} \quad (21)$$

where the first and last equalities are due to (10), 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

$$\begin{aligned} c^\top x_{\bar{k}+1}^*(S) &= c^\top z^*(S \setminus \cup_{j=0}^{\bar{k}} R_j(S)) \\ &= c^\top z^*(\text{supp}(x_{\bar{k}+1}^*(S))) \leq c^\top z^*(C \setminus \cup_{j=0}^{\bar{k}} R_j(S)) \\ &= c^\top x_{\bar{k}+1}^*(C), \end{aligned} \quad (22)$$

where the first and last equalities are due to (10), the second one due to the fact that  $\text{supp}(x_{\bar{k}+1}^*(S)) \subset S \setminus \cup_{j=0}^{\bar{k}} R_j(S)$ , and the inequality holds since  $R_j(C) = R_j(S)$  and  $\text{supp}(x_{\bar{k}+1}^*(S)) \subset C \setminus \cup_{j=0}^{\bar{k}} R_j(S)$ . By (21) and (22)

we then have that  $x_{k+1}^*(C) = x_{k+1}^*(S)$ , thus concluding the induction proof. In other words, we have shown that

$$x_k^*(C) = x_k^*(S), \text{ for all } k \in \{0, \dots, \ell\}. \quad (23)$$

Relation (23) together with the sufficiency part of Proposition 1 shows that the candidate  $C$  in (19) is a compression set.

*Uniqueness:* To show that  $C$  in (19) is the unique compression set, assume for the sake of contradiction that there exists another compression  $C' \subset S$  for the mapping defined in (11),  $C' \neq C$ , with  $|C'| = (\ell+1)d$ . Since  $C' \subset S$  is a compression, Proposition 1 (necessity part) implies that  $x_k^*(C') = x_k^*(S)$ , for all  $k \in \{0, \dots, \ell\}$ , as  $C'$  is a compression. Besides, by the existence part (Step 1 above), we have shown that for  $C$  given in (19) we have that  $x_k^*(C) = x_k^*(S)$  for all  $k \in \{0, \dots, \ell\}$ . We thus have that for all  $k \in \{0, \dots, \ell\}$ ,  $x_k^*(C) = x_k^*(C')$ . This in turn implies that  $\text{supp}(x_k^*(C)) = \text{supp}(x_k^*(C'))$  for all  $k \in \{0, \dots, \ell\}$ , which, by Assumption 2, leads to  $C = C'$  (to see this notice that  $\cup_{k=0}^{\ell} \text{supp}(x_k^*(S)) \subset C'$  and  $|C'| = (\ell+1)d$ ), thus establishing a contradiction.

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

$$\mathcal{A}(C) = (\mathcal{A}_1(C) \cap \mathcal{A}_2(C)) \cup \mathcal{A}_3(C), \quad (24)$$

where the individual sets are as in (11). Recall also that  $\mathcal{A}_3(S)$  is a discrete set that contains the removed samples throughout the execution of the scheme of Figure 1. 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 (19). We have that

$$\begin{aligned} \mathbb{P}\{\mathcal{A}(C)\} &= \mathbb{P}\{(\mathcal{A}_1(C) \cap \mathcal{A}_2(C)) \cup \mathcal{A}_3(C)\} \\ &= \mathbb{P}\{\mathcal{A}_1(C) \cap \mathcal{A}_2(C)\}, \\ &\leq \mathbb{P}\{\mathcal{A}_1(C)\} = \mathbb{P}\{\delta \in \Delta : g(x^*(C), \delta) \leq 0\}, \\ &= \mathbb{P}\{\delta \in \Delta : g(x^*(S), \delta) \leq 0\}, \end{aligned} \quad (25)$$

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  $x^*(C) = x^*(S)$  (see (23)).

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

$$\begin{aligned} \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^*(S), \delta) > 0\} > \epsilon\} \\ \leq \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \notin \mathcal{A}(C)\} > \epsilon\}. \end{aligned} \quad (26)$$

Therefore, since set  $C$  in (19) is the unique compression of  $\mathcal{A}(C)$ , by Theorem 2 we have that

$$\begin{aligned} \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : \delta \notin \mathcal{A}(C)\} > \epsilon\} \\ \leq \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \end{aligned} \quad (27)$$

By (26) and (27) we then have that  $\mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta \in \Delta : g(x^*(S), \delta) > 0\} > \epsilon\} \leq \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}$ , 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 (11) by replacing the second set in its definition with

$$\begin{aligned} \mathcal{A}_2(C) &= \bigcap_{k=0}^{\ell} \{\delta \in \Delta : F_k(z^*(J \cup \{\delta\})) \leq F_k(x_k^*(C))\}, \\ &\text{for all } J \subset C \setminus \cup_{j=0}^{k-1} R_j(C), \text{ with } |J| = d-1, \end{aligned} \quad (28)$$

where  $F_k(\cdot)$  is the augmented objective function defined in (8), related to  $P_k$  defined by means of the regularization procedure of Section III. 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

$$C = \bigcup_{k=0}^{\ell} (\text{supp}(x_k^*(S)) \cup Z_k(S)), \quad (29)$$

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  $|\text{supp}(x_k^*(S))|$  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 (29) is the unique compression set of size  $(\ell+1)d$  for the mapping in (11), with  $\mathcal{A}_2(C)$  in (29) in place of  $\mathcal{A}_2(C)$  in (11). By (7), recall that  $R_k(C) = \text{supp}(x_k^*(C)) \cup Z_k(C)$ ,  $k \in \{0, \dots, \ell-1\}$ .

**Proposition 2.** *Suppose Assumptions 1 and 3 hold. Let  $C$  be the set in (29), and consider the scheme of Figure 1 with the removed scenarios given by (7). We have that:*

- i)  $x_k^*(C) = x_k^*(S)$  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  $x_{\bar{k}}^*(C') \neq x_{\bar{k}}^*(C)$ . Then  $x_{\bar{k}}^*(C') \neq x_{\bar{k}}^*(C' \cup \{\delta\})$  for some  $\delta \in \text{supp}(x_{\bar{k}}^*(C)) \setminus \text{supp}(x_{\bar{k}}^*(C'))$ . 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

$$x_0^*(C) = z^*(C) = z^*(\text{supp}(x_0^*(S))) = x_0^*(S), \quad (30)$$

where the first equality follows from the definition in (10), for the second one we use the definition of the support set, and the third one follows from the definition of  $x_0^*(S)$  and the definition of the support set. Moreover, we have that

$$\begin{aligned} Z_0(C) &= \left\{ \nu_0(S) \text{ scenarios with the smallest labels in} \right. \\ &\quad \left. C \setminus \{\text{supp}(x_0^*(S))\} \right\} \\ &= \left\{ \nu_0(S) \text{ scenarios with the smallest labels in} \right. \\ &\quad \left. S \setminus \{\text{supp}(x_0^*(S))\} \right\} = Z_0(S), \end{aligned} \quad (31)$$

where the first equality is due to the definition of  $C$  in (29) and the fact that  $Z_0(S) \subset C$ , while the last one is due to the definition of  $Z_0(S)$  in (6). Assume now that  $x_k^*(C) = x_k^*(S)$  and  $Z_k^*(C) = Z_k^*(S)$  for all  $k \in \{0, \dots, \bar{k}\}$ , and consider the case  $\bar{k} + 1$ . Indeed, we have that

$$\begin{aligned} x_{\bar{k}+1}^*(C) &= z^*(C \setminus \bigcup_{j=0}^{\bar{k}} R_j(C)) = z^*(\text{supp}(x_{\bar{k}+1}^*(S))) \\ &= z^*(S \setminus \bigcup_{j=0}^{\bar{k}} R_j(S)) = x_{\bar{k}+1}^*(S), \end{aligned} \quad (32)$$

where these relations follow as in (30) for the case  $k = 0$ . We also have that

$$\begin{aligned} Z_{\bar{k}+1}^*(C) &= \left\{ \nu_{\bar{k}+1}(S) \text{ scenarios with the smallest labels in} \right. \\ &\quad \left. C \setminus \left( \bigcup_{j=0}^{\bar{k}} R_j(S) \cup \text{supp}(x_{\bar{k}+1}^*(S)) \right) \right\} = Z_{\bar{k}+1}^*(S), \end{aligned} \quad (33)$$

since  $Z_{\bar{k}+1}^*(S) \subset C \setminus \bigcup_{j=0}^{\bar{k}} \{R_j(S) \cup \text{supp}(x_{\bar{k}+1}^*(S))\}$  due to the particular choice of  $C$  in (29), thus proving that for  $C$  in (29) we have  $x_k^*(C) = x_k^*(S)$  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  $\delta \in \text{supp}(x_k^*(C)) \setminus \text{supp}(x_k^*(C'))$  we have that  $x_k^*(C') = x_k^*(C' \cup \{\delta\})$ . We will show that  $x_k^*(C) = x_k^*(C')$ . We then have that

$$\begin{aligned} c^\top x_k^*(C') &= c^\top x_k^*(C' \cup \{\delta\}) = c^\top x_k^*(C' \cup \text{supp}(x_k^*(C))) \\ &= c^\top x_k^*(C), \end{aligned} \quad (34)$$

where the second equality holds due to Lemma 2.12 in [26] since  $C' \cup \{\delta\} \subset C' \cup \text{supp}(x_k^*(C))$ . 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  $x_k^*(C) = x_k^*(C')$ .

We now show that such a  $\delta$  must belong to  $C \setminus C'$ . In fact, choose  $\bar{\delta} \in \text{supp}(x_k^*(C)) \setminus \text{supp}(x_k^*(C'))$  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

$$\begin{aligned} c^\top x_k^*(C' \cup \{\bar{\delta}\}) &= c^\top z^*\left((C' \setminus \bigcup_{j=0}^{\bar{k}-1} R_j(C')) \cup \{\bar{\delta}\}\right) \\ &= c^\top z^*(\text{supp}(x_k^*(C'))) \\ &= c^\top x_k^*(C') \end{aligned} \quad (35)$$

where the first relation holds due to (10) and the fact that  $R_j(C') = R_j(C)$  for all  $j < \bar{k}$ , the second one is due to the fact that  $\text{supp}(x_k^*(C')) \subset C' \setminus \bigcup_{j=0}^{\bar{k}-1} R_j(C) \cup \{\bar{\delta}\}$  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 (35) contradicts our choice of  $\bar{\delta}$ , which requires that  $x_k^*(C') \neq x_k^*(C' \cup \{\bar{\delta}\})$ . 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  $x_{\bar{k}}^*(C') \neq x_{\bar{k}}^*(C)$ ; or case II:  $x_k^*(C') = x_k^*(C)$  for all  $k \in \{0, \dots, \ell\}$ , but there exists a  $\tilde{k} \in \{0, \dots, \ell\}$  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  $x_{\bar{k}}^*(C') \neq x_{\bar{k}}^*(C)$ , 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

$$\begin{aligned} Z_{\tilde{k}}(C') &= \left\{ \nu_{\tilde{k}}(C) \text{ scenarios with the smallest labels in} \right. \\ &\quad \left. C' \setminus \bigcup_{j=0}^{\tilde{k}-1} R_j(C') \right\} \\ &= \left\{ \nu_{\tilde{k}}(C) \text{ scenarios with the smallest labels in} \right. \\ &\quad \left. C' \setminus \bigcup_{j=0}^{\tilde{k}-1} R_j(S) \right\}, \end{aligned} \quad (36)$$

where the first equality is by the definition in (6) 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  $\delta \in Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(S)$ ,

$$y_\delta > \max_{\xi \in Z_{\tilde{k}}(S)} y_\xi = y_{\max}, \quad (37)$$

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

We will use the relation (37) to show that any element in  $C \setminus C'$  has a label greater than  $y_{\max}$ . In fact, note that

$$C' \setminus C \subset \left\{ \bigcup_{j=\bar{k}+1}^{\ell} R_j(C') \right\} \cup \{Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(C)\}, \quad (38)$$

hence it suffices to show that any element in either set in the right-hand side of (38) is greater than  $y_{\max}$ . To this end, fix any  $\delta \in \bigcup_{j=\bar{k}+1}^{\ell} R_j(C')$  and note that

$$y_\delta > \max_{\xi \in Z_{\tilde{k}}(C') \setminus Z_{\tilde{k}}(C)} y_\xi > y_{\max}, \quad (39)$$

where the first inequality is due to the fact that since such a  $\delta$  has not been removed up to stage  $\bar{k}$ , then its label will be

greater than the ones in  $Z_{\bar{k}}(C')$ , and as a result the ones in  $Z_{\bar{k}}(C') \setminus Z_{\bar{k}}(C)$ . The second inequality follows from (37) and the fact that  $Z_{\bar{k}}(C') \setminus Z_{\bar{k}}(C) \subset Z_{\bar{k}}(S)$ . Therefore, for any  $\delta \in C' \setminus C$  we have that  $y_{\bar{\delta}} > y_{\max}$ .

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

$$F_{\bar{k}}(z^*(\bar{J} \cup \{\bar{\delta}\})) > F_{\bar{k}}(x_k^*(C)), \quad (40)$$

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 (40).

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  $\bar{\delta} \in \{\text{supp}(x_k^*(C)) \setminus \text{supp}(x_k^*(C'))\} \cap \{C \setminus C'\}$  such that  $x_k^*(C') \neq x_k^*(C' \cup \{\bar{\delta}\})$ . Repeating the arguments following equations (14) and (15) 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 (29).

*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 (11), with the set  $\mathcal{A}_2(C)$  in (11) being substituted with the one in (28). The arguments then follows *mutatis mutandis* the ones in the last part of the fully-supported case. This concludes the proof of Theorem 4.  $\square$

## V. 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 1 achieves the upper bound given by the right-hand side of (9) when all the intermediate problems  $P_k, k = 0, \dots, \ell$ , 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 (11) with  $\bar{\mathcal{A}} : \Delta^m \rightarrow 2^\Delta$  defined

$$\bar{\mathcal{A}}(C) = \left\{ \delta \in \Delta : g(x_\ell^*(C), \delta) \leq 0 \right\} \cup \left\{ \bigcup_{k=0}^{\ell-1} \text{supp}(x_k^*(C)) \right\}. \quad (41)$$

Note that  $\bar{\mathcal{A}}(C)$  coincides with the one in (24), but without the set  $\mathcal{A}_2(C)$  in its definition. We impose the following assumption.

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

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

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

Assumption 4 imposes a restriction on the class of fully-supported problems. For instance, the pictorial example of Figure 2 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, whenever a sample belongs to the support scenarios of any intermediate problem, then the scenario associated with it is violated by all the solutions that could have been obtained using any subset of cardinality  $d$  from the remaining samples. Note that verifying Assumption 4 is hard in general; we show in the next subsection an example that satisfies this requirement and admits an analytic solution. Assumption 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 1 to show that the result of Theorem 3 is tight. This serves as a constructive argument for the existential result of [25].

Note that in this paper we do not offer any means to check the validity of Assumption 4; however, we show that this class of problems is not empty in the next section.

**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 (4) and illustrated in Figure 1, and let the minimizer of the  $\ell$ -th program be  $x^*(S) = x_\ell^*(S)$ . We then have that

$$\mathbb{P}^m \left\{ (\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P} \{ \delta \in \Delta : g(x^*(S), \delta) > 0 \} > \epsilon \right\} = \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}. \quad (42)$$

*Proof. Existence:* We first show that the set  $C$  given in (19) is a compression for the mapping in (41). Recall that under Assumption 2 we have that  $R_k(S) = \text{supp}(x_k^*(S))$  for all  $k \in \{0, \dots, \ell\}$ . Applying a similar induction argument as in the existence part of Theorem 3, we have that  $x_k^*(C) = x_k^*(S)$  for all  $k \in \{0, \dots, \ell\}$ . Hence, by the definition of the mapping  $\bar{\mathcal{A}}(C)$  in (41), we obtain that  $\bar{\mathcal{A}}(C) = \bar{\mathcal{A}}(S)$ , thus showing that  $C$  in (19) is a compression.

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

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

$$x_k^*(C') = x_k^*(C' \cup \{\delta\}), \text{ for all } k \in \{0, \dots, \ell\}. \quad (43)$$

In fact, if (43) holds for all  $\delta \in S \setminus C'$  by induction it follows then that  $x_k^*(C') = x_k^*(S)$  for all  $k \in \{0, \dots, \ell\}$ .

To show (43) 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  $x_k^*(C') \neq x_k^*(C' \cup \{\bar{\delta}\})$ . Let  $\bar{k}$  be the smallest index such that this occurs and note that

$$x_{\bar{k}}^*(C') = z^*(C' \setminus \bigcup_{j=0}^{\bar{k}-1} \text{supp}(x_j^*(C'))), \quad (44)$$

$$x_{\bar{k}}^*(C' \cup \{\bar{\delta}\}) = z^*((C' \setminus \bigcup_{j=0}^{\bar{k}-1} \text{supp}(x_j^*(C')) \cup \{\bar{\delta}\}), \quad (45)$$

which implies that  $\bar{\delta} \in \text{supp}(x_{\bar{k}}^*(C' \cup \{\bar{\delta}\}))$ , as removal of  $\bar{\delta}$  will change  $x_{\bar{k}}^*(C' \cup \{\bar{\delta}\})$  to  $x_{\bar{k}}^*(C')$ . By Assumption 4 and since  $\text{supp}(x_j^*(C')) = \text{supp}(x_j^*(C' \cup \{\bar{\delta}\}))$  for all  $j = 0, \dots, \bar{k} - 1$ , we have that for all  $J \subset C' \setminus (\bigcup_{j=0}^{\bar{k}-1} \text{supp}(x_j^*(C')) \cup \{\bar{\delta}\})$  with cardinality  $d$ ,

$$g(z(J), \bar{\delta}) > 0. \quad (46)$$

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

$$g(z(\bar{J}), \bar{\delta}) = g(x_{\bar{k}}^*(C'), \bar{\delta}) > 0, \quad (47)$$

where the equality follows from (10). However,  $C'$  is assumed to be a compression set for  $\bar{A}$ , which implies that  $\delta \in \bar{A}(C')$ , i.e.,  $g(x_{\bar{k}}^*(C'), \bar{\delta}) \leq 0$ . This is in contradiction with (47), implying that  $x_k^*(C') = x_k^*(C' \cup \delta)$ , 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  $x_k^*(C') = x_k^*(S) = x_k^*(C)$  for all  $k \in \{0, \dots, \ell\}$ , thus showing that  $C$  in (19) is the unique compression set for the mapping defined in (41).

By Theorem 2, we then have that

$$\begin{aligned} & \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta : \delta \notin \bar{A}(C)\} > \epsilon\} \\ &= \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta : g(x_{\bar{k}}^*(C), \delta) > 0\} > \epsilon\} \\ &= \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \mathbb{P}\{\delta : g(x_{\bar{k}}^*(S), \delta) > 0\} > \epsilon\} \\ &= \sum_{i=0}^{r+d-1} \binom{m}{i} \epsilon^i (1-\epsilon)^{m-i}, \end{aligned} \quad (48)$$

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  $x_{\bar{k}}^*(C) = x_{\bar{k}}^*(S)$  for the compression set defined in (19). This concludes the proof of Theorem 5.  $\square$

## B. An example with an analytic solution

We revisit the simple problem studied in [25], [34] and show that it satisfies Assumption 4. We compute analytically the violation probability of the solution returned by applying the scheme of Figure 1 to this problem and show that the resulting violation probability is in line with the result of Theorem 5.

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

$$\begin{aligned} & \text{minimize}_{x \in [0,1]} \quad x \\ & \text{subject to} \quad x \geq \delta_i, \quad i \in S \setminus \bigcup_{j=0}^{k-1} R_j(S). \end{aligned} \quad (49)$$

We further assume that all samples are extracted from a uniform distribution over the interval  $[0, 1]$ . Note that (49) satisfies Assumptions 1 and 2. Also notice that Assumption 4 is satisfied for this problem, as  $x_k^* = \max_{i \in S \setminus \bigcup_{j=0}^{k-1} R_j(S)} \delta_i$ , i.e., the maximum among the scenarios available at stage  $k$  of the discarding process. Under the choice of a uniform distribution, the support set is a singleton, i.e., the maximizing scenario is unique, with  $\mathbb{P}^m$ -probability one. Therefore, once the single support scenario is removed, the new minimizer will necessarily be lower thus violating the removed scenario. Note that since this is an one-dimensional problem ( $d = 1$ ), our procedure involves removing samples one by one.

Let  $\epsilon \in (0, 1)$  and  $r = \ell d = \ell < m$  (since this is an one-dimensional problem), and consider the sets:

$$B = \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : x^*(S) < 1 - \epsilon\}, \quad (50)$$

which represents the measure of samples such that the probability of constraint violation is greater than  $\epsilon$ , as  $\mathbb{P}\{\delta \in \Delta : x^*(S) < \delta\} = 1 - x^*(S)$  due to the fact that  $\mathbb{P}$  is uniform on  $[0, 1]$ . Moreover,

$$\begin{aligned} A_0 &= \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \text{for all } i = 1, \dots, m, \\ & \quad \delta_i \leq 1 - \epsilon\}, \end{aligned} \quad (51)$$

and for all  $\ell = 1, \dots, m$ ,

$$A_i = \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \text{there exist exactly } i \text{ samples greater than } 1 - \epsilon\}. \quad (52)$$

Observe that  $\{A_i\}_{i=0}^m$  forms a partition of  $\Delta^m$ , and notice that

$$\mathbb{P}^m\{A_0\} = \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : \delta_i \leq 1 - \epsilon, \text{ for all } i = 1, \dots, m\} = (1 - \epsilon)^m, \quad (53)$$

where the last equality is due to sample independence. Since  $\mathbb{P}^m\{\delta_i \in \Delta : \delta_i > 1 - \epsilon\} = \epsilon$  (due to the fact that  $\mathbb{P}$  is uniform on  $[0, 1]$ ) and  $A_i$  involves  $i$  independent samples, we have that

$$\mathbb{P}^m\{A_i\} = \binom{m}{i} \epsilon^i, \quad (54)$$

where the factor  $\binom{m}{i}$  accounts for all combinations of  $i$  out of  $m$  samples. Moreover, for  $i \leq r$ ,  $\mathbb{P}^m\{B|A_i\} = (1 - \epsilon)^{m-i}$  since it involves conditioning on exactly  $i$  samples being

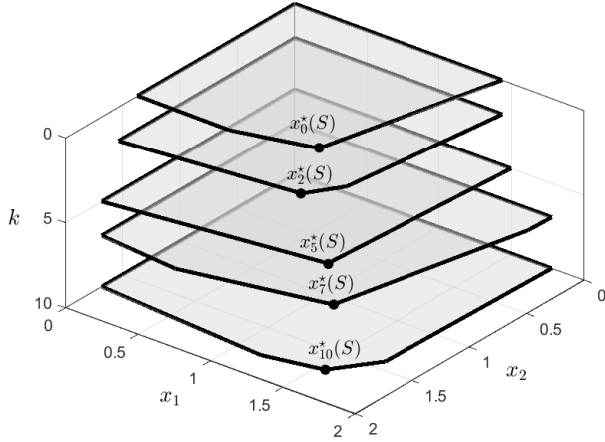


Fig. 4: Feasibility sets of the intermediate problems  $P_k$ ,  $k = 0, 2, 5, 7, 10$ , for the scheme proposed in Figure 1 when applied to (57). The optimal solution of each problem is denoted by  $x_k^*(S)$ ,  $k = 0, 2, 5, 7, 10$ .

removed, and computing the probability that the returned solution  $x^*(S)$  is feasible for the remaining  $m - i$  samples. By the total law of probability, we then have that

$$\begin{aligned} \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : x^*(S) < 1 - \epsilon\} &= \mathbb{P}^m\{B\} \\ &= \sum_{i=0}^m \mathbb{P}^m\{A_i \cap B\} \\ &= \sum_{i=0}^r \mathbb{P}^m\{A_i \cap B\} + \sum_{i=r+1}^m \mathbb{P}^m\{A_i \cap B\}. \end{aligned} \quad (55)$$

The terms involving  $i > r$  are zero because  $(\delta_1, \dots, \delta_m) \in B$  implies that at most  $r$  of samples are greater than  $1 - \epsilon$ , i.e.,  $\mathbb{P}^m\{B|A_i\} = 0$ , for all  $i > r$ . Hence,

$$\begin{aligned} \mathbb{P}^m\{(\delta_1, \dots, \delta_m) \in \Delta^m : x^*(S) < 1 - \epsilon\} &= \sum_{i=0}^r \mathbb{P}^m\{A_i \cap B\} \\ &= \sum_{i=0}^r \mathbb{P}^m\{B|A_i\} \mathbb{P}^m\{A_i\} = \sum_{i=0}^r \binom{m}{i} \epsilon^i (1 - \epsilon)^{m-i}, \end{aligned} \quad (56)$$

where the last equality follows from (53) and (54). Note that this result coincides with the bound of Theorem 5.

## VI. 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  $p = 1, \dots, n$  and for all  $j = 1, \dots, d$ , and solve the following convex scenario problem

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && A(\delta_i)x \leq b, \quad \text{for all } i = 1, \dots, m, \end{aligned} \quad (57)$$

where, for each  $i \in \{1, \dots, m\}$ ,  $A(\delta_i) \in \mathbb{R}^{n \times d}$  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  $c = [-1 \ \dots \ -1]^\top \in \mathbb{R}^d$ .

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

Figure 4 illustrates the feasible set for stages  $k = 0, 2, 5, 7$ , and 10 of the scheme of Figure 1, and depicts the corresponding optimal solution for each  $P_k$  as  $x_k^*(S)$ . 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 5 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, using numerical inversion and  $\beta = 1 \times 10^{-6}$ .

Consider now (57) 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 in 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

<sup>1</sup>For our simulations, fix  $i \in \{1, \dots, m\}$  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.

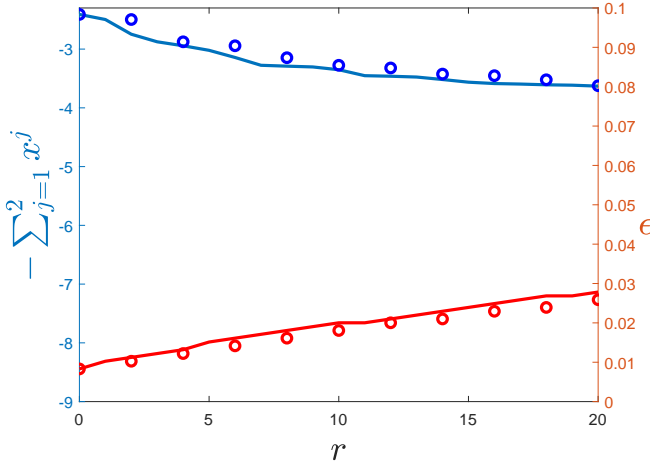


Fig. 5: Cost and probability of constraint violation for the solution returned by the scheme of Figure 1 and a greedy removal strategy for the problem in (57) when  $d = 2$ ,  $n = 2$ , and  $m = 2000$ . With the blue dots we show the cost obtained by the proposed procedure, where we are allowed to remove 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, using numerical inversion and  $\beta = 1 \times 10^{-6}$ .

can be removed using the bound in (4) is  $r = 18$ , but we only remove 10. Figure 6 shows then the relative cost difference  $100 \times \frac{f^*(\epsilon) - \bar{f}^*(\epsilon)}{f^*(\epsilon)}$  as a function of  $\epsilon$ , where  $f^*(\epsilon)$  is the optimal value of problem (57) when scenarios are removed according to Theorem 4, and  $\bar{f}^*(\epsilon)$  correspond to the bound in [25]. Note that for  $\epsilon \geq 0.03$  our scheme results in an improvement in performance 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 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.

Note that the computational requirements of the proposed approach are lower with respect to the greedy removal strategy of [25] (see also [26]). To put this in perspective, to remove 100 scenarios in the previous example when  $d = 10$ , the greedy strategy, as described in [25], requires the solution of 1101 optimization problems of the form (57), 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 3.** Note that Theorem 4 offers an improvement with respect to Theorem 1, [25], [26], as far as probabilistic guarantees on the probability of constraint violation are concerned. However, the effect of the proposed discarding scheme with respect to the greedy removal strategy of [25], [26], does

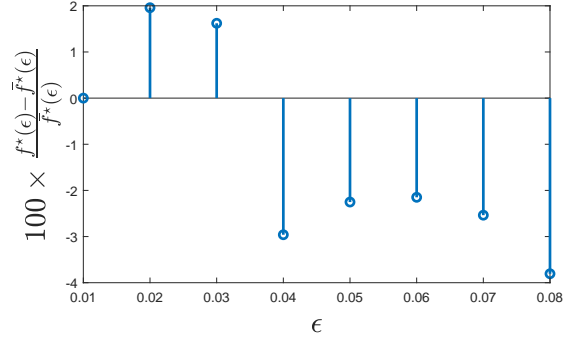


Fig. 6: Relative cost improvement  $100 \times \frac{f^*(\epsilon) - \bar{f}^*(\epsilon)}{f^*(\epsilon)}$ , as a function of  $\epsilon$ , where  $f^*(\epsilon)$  corresponds to the cost associated with Theorem 4, and  $\bar{f}^*(\epsilon)$  to the one of Theorem 1, [25]. The numerical results correspond to (57) with  $d = 10$ .

not necessarily follow the one of Figure 6, but is problem dependent.

## VII. 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 tighter bound on the probability of constraint violation of the obtained solution, extending state-of-the-art bounds. 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 an example that admits an analytic solution, and on a resource sharing linear program.

Current work concentrates towards extending our scenario discarding scheme so that we no longer remove scenarios in batches but one by one. 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.

## ACKNOWLEDGEMENT

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

## REFERENCES

- [1] M. Vidyasagar, "Statistical Randomized Learning Theory and Algorithms for Control," *IEEE Control Systems Magazine*, vol. 18, no. 6, pp. 69–85, 1998.
- [2] T. Alamo, R. Tempo, and E. F. 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. C. Calafiore, "Repetitive Scenario Design," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1125–1137, 2017.
- [6] F. A. Ramponi, "Consistency of the scenario approach," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 135–162, 2018.
- [7] M. C. Campi, S. Garatti, and F. A. 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. C. 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. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," 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. E. Ghaoui, and A. Nemirovski, *Robust optimization*, 2009.
- [13] M. C. De Oliveira, J. Bernussou, and J. C. Geromel, "A new discrete-time robust stability condition," *Systems and Control Letters*, vol. 37, no. 4, pp. 261–265, 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. C. Geromel, M. C. 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. C. 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. B. Romao, M. C. De Oliveira, P. L. Peres, and R. C. Oliveira, "State-feedback and filtering problems using the generalized KYP lemma," in *Proceedings of the IEEE International Symposium on Computer-Aided Control System Design*, vol. 2016-October, 2016, pp. 1054–1059.
- [19] H. W. Kuhn and A. Prekopa, "On Probabilistic Constrained Programming," in *Proceedings of the Princeton Symposium on Mathematical Programming*, 1970, 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. K. 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. C. Campi, "Uncertain convex programs: Randomized solutions and confidence levels," *Mathematical Programming*, vol. 102, no. 1, pp. 25–46, 2005.
- [23] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, 2006.
- [24] M. C. 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] M. Campi and S. Garatti, "A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality," *J Optim Theory Appl*, vol. 148, pp. 257–280, 2011.
- [26] G. C. 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. C. Campi, "Complexity-based modulation of the dataset in scenario optimization," *European Control Conference*, no. 2, pp. 1386–1391, 2019.
- [30] M. Vidyasagar, *Learning and Generalization*. Springer, 2002.
- [31] S. Floyd and M. Warmuth, "Sample compression, learnability, and the Vapnik-Chervonenkis dimension," *Machine Learning*, no. 21, pp. 269–304, 1995.
- [32] S. Garatti and M. C. Campi, "Risk and complexity in scenario optimization," *Mathematical Programming*, pp. 1–37, 2019.
- [33] D. A. Salamon, *Measure and Integration*. Zurich, Switzerland: European Mathematical Society, 2016, 2016.
- [34] M. C. Campi and G. C. Calafiore, "Notes on the scenario design approach," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 382–385, 2009.



