Chance-constrained sets approximation: A probabilistic scaling approach

Chance-constrained sets approximation: A probabilistic scaling approach

M. Mammarella [a] V. Mirasierra [b] M. Lorenzen [c] T. Alamo [b] F. Dabbene [a]

[a] *CNR-IEIIT; c/o Politecnico di Torino; Corso Duca degli Abruzzi 24, Torino; Italy.*

[b] *Universidad de Sevilla, Escuela Superior de Ingenieros, Camino de los Descubrimientos s/n, Sevilla; Spain.*

[c] *Systemwissenschaften, TTI GmbH, Nobelstr. 15, 70569 Stuttgart, Germany.*

Abstract

In this paper, a sample-based procedure for obtaining simple and computable approximations of chance-constrained sets is proposed. The procedure allows to control the complexity of the approximating set, by defining families of simple-approximating sets of given complexity. A probabilistic scaling procedure then scales these sets to obtain the desired probabilistic guarantees. The proposed approach is shown to be applicable in several problems in systems and control, such as the design of Stochastic Model Predictive Control schemes or the solution of probabilistic set membership estimation problems.

1 Introduction

In real-world applications, the complexity of the phenomena encountered and the random nature of data makes dealing with uncertainty essential. In many cases, uncertainty arises in the modeling phase, in some others it is intrinsic to both the system and the operative environment, as for instance wind speed and turbulence in aircraft or wind turbine control [1]. Deriving results in the presence of uncertainty is of major relevance in different areas, including, but not limited to, optimization [2] and robustness analysis [3]. However, with respect to robust approaches, where the goal is to determine a feasible solution which is optimal in some sense for *all* possible uncertainty instances, the goal in the stochastic framework is to find a solution that is feasible for *almost* all possible uncertainty realizations, [4, 5]. In several applications, including engineering and finance, uncertainties in price, demand, supply, currency exchange rate, recycle and feed rate, and demographic condition are common. In these situations, it is acceptable, up to a certain safe level, to relax the inherent conservativeness of robust constraints enforcing probabilistic constraints. More recently, this probabilistic approach has been used also in unmanned autonomous vehicle navigation [6, 7] as well as in optimal power flow [8, 9].

In the optimization framework, constraints involving stochastic parameters that are required to be satisfied with a pre-specified probability threshold are called *chance constraints* (CC). In general, dealing with CC implies facing two serious challenges: the solution of difficult parameterized probability integrals and the nonconvexity of the ensuing constraints [10]. Consequently, while being attractive from a modeling viewpoint, problems involving CC are often computationally intractable, generally shown to be NP-hard, which seriously limits their applicability. However, being able to efficiently solve or approximate chance-constrained problems remains an important challenge, especially in systems and control. In the case of approximated solutions, there exists of course a fundamental trade-off between complexity of the approach and goodness of the approximation.

The scientific community has devoted large research in devising computationally efficient approaches to deal with chance-constraints. We review such techniques in Section 3, where we highlight three mainstream approaches: i) *exact techniques*; ii) *robust approximations*; and iii) *sample-based approximations.*

⋆

This research was partially funded by the project PRIN 2017, Prot. 2017S559BB, and has received support by the Agencia Estatal de Investigación (AEI)-Spain under Grant PID2019-106212RB-C41/AEI/10.13039/501100011033.

Email addresses: martina.mammarella@ieiit.cnr.it (M. Mammarella), vmirasierra@us.es (V. Mirasierra), lorenzen@ist.uni-stuttgart.de (M. Lorenzen), talamo@us.es (T. Alamo), fabrizio.dabbene@ieiit.cnr.it (F. Dabbene).

In this paper, we present what we consider an important step forward in the sample-based approach. More precisely, our developments stem from the observation that, while in the general situation one is interested in finding an optimal solution to a chance-constrained problem, there exists a significant class of practical applications in which, instead, what it is really needed is being able to *construct good approximation of the chance-constrained set*. This is the case, for instance, of stochastic model predictive control (SMPC), where this approximation is necessary for post-processing in real time, see for instance [11, 12].

Motivated by these considerations, we propose a simple and efficient strategy to obtain a probabilistically guaranteed inner approximation of a chance-constrained set, with given confidence.

In particular, we describe a two-step procedure that involves: i) the preliminary approximation of the chance-constrained set by means of a so-called Simple Approximating Set (SAS); ii) a sample-used scaling procedure that allows to properly scale the SAS so to guarantee the desired probabilistic properties. The proper selection of a low-complexity SAS allows the designer to easily tune the complexity of the approximating set, significantly reducing the sample complexity. We propose several candidate SAS shapes, grouped in two classes: i) sampled-polytopes; and ii) norm-based SAS.

The approach we propose distinguishes itself from the previous literature on CC in the following main points.

(1) It is specifically tailored towards the specific problem of *approximating the chance-constrained set*, as opposed to solving a specific instance of a chance-constrained problem.
(2) It is very *general*: it applies to a very general class of uncertainty configurations. A large part of the methods available in the literature are limited to cases where the constraints depend in a "nice" way on the uncertainty. This is the case for instance of the solution proposed in [13, 14]. The reader is referred to Section 3 for an overview.
(3) It is highly *tunable*: by selecting the complexity of the approximating set, the designer has a very efficient way to control the trade-off between computational complexity and potential goodness of the approximation.

The probabilistic scaling approach was presented in the conference papers [15,16] and it is based on recent results on order statistics [17]. The present work extends [15,16] in several directions. First, we perform here a thorough mathematical analysis of probabilistic scaling. Second, we provide probabilistic guarantees for a more general class of norm-based SAS. Third, we consider here *joint chance constraints*. This choice is motivated by the fact that enforcing joint chance constraints, which have to be satisfied simultaneously, adheres better to some applications, despite the inherent complexity. Finally, we present here a second application, besides SMPC, related to probabilistic set-membership identification.

The paper is structured as follows. Section 2 provides a general preamble of the problem formulation and of chance-constrained optimization, including two motivating examples. An extensive overview on methods for approximating chance-constrained sets is reported in Section 3 whereas the probabilistic scaling approach has been detailed in Section 4. Section 5 and Section 6 are dedicated to the definition of selected candidate SAS, i.e. sampled-polytope and norm-based SAS, respectively. Last, in Section 7, we validate the proposed approach with a numerical example applying our method to a probabilistic set membership estimation problem. Main conclusions and future research directions are addressed in Section 8.

Notation

Given an integer N, $[N]$ denotes the integers from 1 to N. Given $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the greatest integer no larger than x. Given the ℓ_p-norm $\|\cdot\|_p$, we denote with $\mathbb{B}_p^s$ the ℓ_p-norm ball of radius one in $\mathbb{R}^s$, i.e. $\mathbb{B}_p^s \doteq \{\, z \in \mathbb{R}^s \;:\; \|z\|_p \leq 1 \,\}$. The Chebyshev center of a given set $\mathbb{X}$, with respect to norm $\|\cdot\|_p$, is denoted as $\mathsf{Cheb}_p(\mathbb{X})$, and it is defined as the center of the largest ℓ_p-norm ball inscribed in $\mathbb{X}$. Given an ℓ_p-norm $\|\cdot\|_p$, its dual norm $\|\cdot\|_{p^*}$ is defined as $\|c\|_{p^*} \doteq \sup_{z \in \mathbb{B}_p^s} c^\top z,\ \forall c \in \mathbb{R}^s$. In particular, the couples (p, p^*): $(2,2)$, $(1,\infty)$, $(\infty,1)$ give rise to dual norms [18, A.1.6].
Given two sets $\mathbb{S}_1$ and $\mathbb{S}_2$, the notation $\mathbb{S}_1 \oplus \mathbb{S}_2$ ($\mathbb{S}_1 \ominus \mathbb{S}_2$) stands for the Minkovski sum (difference) between the two sets. Given integers k, N, and parameter $\varepsilon \in [0,1]$, the Binomial cumulative distribution function is denoted as

$$\mathbf{B}(k; N, \varepsilon) \doteq \sum_{i=0}^{k} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}.$$

The following definition is borrowed from the field of order statistics [17, 19].

Definition 1 (Generalized Min) *Given a collection of N scalars $\Gamma = \{\gamma_i\}_{i=1}^N \in \mathbb{R}^N$, and an integer $r \in [N]$, we say that $\gamma_r^- \in \Gamma$ is the r-smallest value of Γ if there is no more than $r-1$ elements of Γ strictly smaller than γ_r^-.*

Hence, γ_1^- denotes the smallest value in Γ, γ_2^- the second smallest one, and so on until γ_N^-, which is equal to the largest one. We also use the alternative notation

$$\min^{(r)}(\Gamma) \doteq \gamma_r^-.$$

2 Problem formulation

Consider a robustness problem, in which the controller parameters and auxiliary variables are parametrized by means of a decision variable vector $\theta \in \mathbb{R}^{n_\theta}$, which is usually referred to as *design parameter*.

Furthermore, the uncertainty vector $w \in \mathbb{R}^{n_w}$ represents one of the admissible uncertainty realizations of a random vector with given probability distribution $\mathsf{Pr}_{\mathbb{W}}$ and (possibly unbounded) support $\mathbb{W}$.

This paper deals with the special case where the design specifications can be decoded as a set of n_ℓ uncertain linear inequalities

$$F(w)\theta \leq g(w), \tag{1}$$

where

$$F(w) = \begin{bmatrix} f_1^\top(w) \\ \vdots \\ f_{n_\ell}^\top(w) \end{bmatrix} \in \mathbb{R}^{n_\ell \times n_\theta}, \quad g(w) = \begin{bmatrix} g_1(w) \\ \vdots \\ g_{n_\ell}(w) \end{bmatrix} \in \mathbb{R}^{n_\ell},$$

are measurable functions of the uncertainty vector $w \in \mathbb{R}^{n_w}$.

In Section 8 we discuss possible extensions of this approach to more general settings, in which the constraints may be nonlinear and even nonconvex. Also note that the proposed setup captures the special case of chance constraints with random right-hand side. These correspond to the choice $F(w) = F$ and $g(w) = w$. Similarly, the case of chance constraints with random technology matrix is captured by our general case.

We also note that hard linear constraints on θ may be directly incorporated by introducing *deterministic* inequalities of the form $f_\ell^\top \theta \leq g_\ell$, where f_ℓ and g_ℓ do not depend on the uncertainty w.

The inequality in (1) is to be interpreted componentwise, i.e. $f_\ell^\top(w)\theta \leq g_\ell(w)$, $\forall \ell \in [n_\ell]$. Due to the random nature of the uncertainty vector w, each realization of w corresponds to a different set of linear inequalities. Consequently, each value of w gives raise to a corresponding set

$$\mathbb{X}(w) \doteq \{\, \theta \in \mathbb{R}^{n_\theta} \;:\; F(w)\theta \leq g(w) \,\}. \tag{2}$$

In every application, one usually accepts a risk of violating the constraints. This often translates into a two-step strategy: i) a set $\tilde{\mathbb{W}}$ is obtained such that $w \in \tilde{\mathbb{W}}$ is satisfied with a pre-specified high probability; ii) a robust design problem in which inequality (1) is forced to be satisfied for every $w \in \tilde{\mathbb{W}}$ is solved. This is for instance the approach proposed in [13, 20]. This approach suffers from several drawbacks: i) there is not guarantee that the ensuing robust problem is easily solvable. Indeed, this may be in general very hard, and to obtain computable solutions the authors of [13] need to make additional assumptions on the dependence of F, g on the uncertainty w. ii) the approach in [13] does not provide a safe region (i.e. a probabilistic approximation of the chance-constrained set), but just a point satisfying the probabilistic constraint. If one uses this approach, then the result may be conservative due to the two-step procedure.

In this paper, we observe that a less conservative solution can be found by choosing the set $\mathbb{W}$ to encompass all possible values and characterizing the region of the design space in which the fraction of elements of $\mathbb{W}$, that violate the constraints, is below a specified level. This concept is rigorously formalized by means of the notion of *probability of violation* (see [5]).

Definition 2 (Probability of violation) *Consider a probability measure $\Pr_{\mathbb{W}}$ over $\mathbb{W}$ and let $\theta \in \mathbb{R}^{n_\theta}$ be given. The probability of violation of θ relative to inequality* (1) *is defined as*

$$\mathsf{Viol}(\theta) \doteq \mathsf{Pr}_{\mathbb{W}}\{\, F(w)\theta \not\leq g(w) \,\}.$$

Given a constraint on the probability of violation, i.e. $\mathsf{Viol}(\theta) \leq \varepsilon$, we denote as (joint) *chance-constrained set* of probability ε (shortly, ε-CCS) the region of the design space for which this probabilistic constraint is satisfied. This is formally stated in the next definition.

Definition 3 (ε-CCS) *Given $\varepsilon \in (0,1)$, we define the chance-constrained set of probability ε as follows*

$$\mathbb{X}_\varepsilon \doteq \{\, \theta \in \mathbb{R}^{n_\theta} \;:\; \mathsf{Viol}(\theta) \leq \varepsilon \,\}. \tag{3}$$

Note that the ε-CCS represents the region of the design space $\mathbb{R}^{n_\theta}$ for which this probabilistic constraint is satisfied and it is equivalently defined as

$$\mathbb{X}_\varepsilon \doteq \Big\{\theta \in \mathbb{R}^{n_\theta} \;:\; \mathsf{Pr}_{\mathbb{W}}\{F(w)\theta \leq g(w)\} \geq 1-\varepsilon\Big\}. \tag{4}$$

Remark 1 (Joint vs. individual CC) *The chance-constraint $\theta \in \mathbb{X}_\varepsilon$, with $\mathbb{X}_\varepsilon$ defined in* (4)*, describes a joint chance constraint. That is, it requires that the joint probability of satisfying the inequality constraint $F(w)\theta \leq g(w)$ is guaranteed to be no smaller than the probabilistic level $1-\varepsilon$. We remark that this constraint is notably harder to impose than individual CC, i.e. constraints of the form*

$$\theta \in \mathbb{X}_{\varepsilon_\ell}^\ell \doteq \Big\{\theta \in \mathbb{R}^{n_\theta} \;:\; \mathsf{Pr}_{\mathbb{W}}\left\{f_\ell(w)^\top\theta \leq g_\ell(w)\right\} \geq 1-\varepsilon_\ell\Big\}, \quad \ell \in [n_\ell],$$

with $\varepsilon_\ell \in (0,1)$. A discussion on the differences and implications of joint and individual chance constraints may

be found in several papers, see for instance [10, 21] and references therein. Note that a well-known (conservative) approximation to the joint chance-constrained set is the use of multiple individual CC.

While there exist simple examples for which a closed-form evaluation of $\mathbb{X}_\varepsilon$ is possible, see e.g. [15, Figure 1], we remark that this is not the case in general. Indeed, as pointed out in [10], typically the computation of the ε-CCS is extremely difficult, since the evaluation of the probability $\mathsf{Viol}(\theta)$ amounts to the solution of a multivariate integral, which is NP-Hard [22].
Moreover, the set ε-CCS is often nonconvex, except for very special cases. For example, [23, Lemma 4.60] shows that the solution set of separable chance constraints can be written as the union of cones, which is nonconvex in general.

Example 1 (Example of nonconvex ε-CCS) *To illustrate these inherent difficulties, we consider the following three-dimensional example ($n_\theta = 3$) with $w = \{w_1, w_2\}$, where the first uncertainty $w_1 \in \mathbb{R}^3$ is a three-dimensional normal-distributed random vector with zero mean and covariance matrix*

$$\Sigma = \begin{bmatrix} 4.5 & 2.26 & 1.4 \\ 2.26 & 3.58 & 1.94 \\ 1.4 & 1.94 & 2.19 \end{bmatrix},$$

and the second uncertainty $w_2 \in \mathbb{R}^3$ is a three-dimensional random vector whose elements are uniformly distributed in the interval $[0, 1]$. The set of viable design parameters is given by $n_\ell = 4$ uncertain linear inequalities of the form

$$F(w)\theta \leq \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^\top, \qquad (5)$$
$$F(w) = \begin{bmatrix} w_1 & w_2 & (2w_1 - w_2) & w_1^2 \end{bmatrix}^\top,$$

where the square power w_1^2 is to be interpreted element-wise.
In this case, to obtain a graphical representation of the set $\mathbb{X}_\varepsilon$, we resorted to gridding the design set, and, for each point θ in the grid, to approximate the probability through a Monte Carlo method. This procedure is clearly unaffordable for higher dimensions. In Figure 1 we report the plot of the obtained ε-CCS set for different values of ε. We observe that the set is indeed nonconvex.

2.1 Chance-constrained optimization

Finding an optimal $\theta \in \mathbb{X}_\varepsilon$ for a given cost function $J : \mathbb{R}^{n_\theta} \to \mathbb{R}$, leads to the *chance-constrained optimization* (CCO) problem

$$\min_{\theta \in \mathbb{X}_\varepsilon} J(\theta), \qquad (6)$$

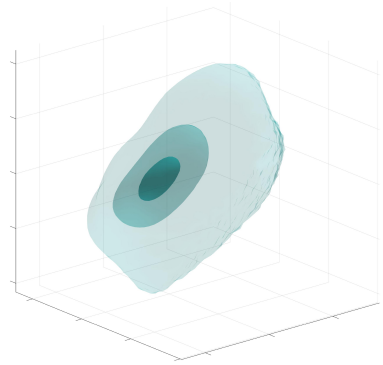


Fig. 1. The ε-CCS set for $\varepsilon = 0.15$ (smaller set), $\varepsilon = 0.30$ (intermediate set), and $\varepsilon = 0.45$ (larger set). We observe that all sets are nonconvex, but the nonconvexity is more evident for larger values of ε, corresponding to larger levels of accepted violation, while the set $\mathbb{X}_\varepsilon$ appears "almost convex" for small values of ε. This kind of behaviour is in accordance with a recent result that proves convexity of the ε-CCS for small enough values of ε, and it is usually referred to as *eventual convexity* [24], [25].

where the cost-function $J(\theta)$ is usually assumed to be a convex, often even quadratic or linear, function.

We remark that the solution of the CCO problem (6) is in general NP-hard, for the same reasons reported before. We also note that several stochastic optimization problems arising in different application contexts can be formulated as a CCO. Typical examples are for instance the reservoir system design problem proposed in [26], where the problem is to minimize the total building and penalty costs while satisfying demands for all sites and all periods with a given probability, or the cash matching problem [27], where one aims at maximizing the portfolio value at the end of the planning horizon while covering all scheduled payments with a prescribed probability. CCO problems also frequently arise in short-term planning problems in power systems. These optimal power flow (OPF) problems are routinely solved as part of the real-time operation of the power grid. The aim is determining minimum-cost production levels of controllable generators subject to reliably delivering electricity to customers across a large geographical area, see e.g. [8] and references therein.

Recently, approaches based on a probabilistic approximation of chance-constrained sets have emerged in the context of Stochastic MPC, see [11, 12, 28]. These approaches exploit the sample-based results we summarize in Section 3.3 to construct *offline* a probabilistically guaranteed approximation of the set of all couples of control input/initial states that guarantee fulfillment of the desired input/state constraints. The possibility of constructing the approximation offline constitutes a winning feature with respect to similar approaches based on samples, since it moves all the cumbersome computation to the control design phase. In the online implementation,

the only operation to be performed is to "evaluate" this set in correspondence to the current initial state. In this way, the original stochastic optimization program is reduced to an efficiently solvable quadratic program. This represents an undiscussed advantage, which has been demonstrated for instance in [12]. We stress that the key element of this procedure is exactly the construction of a "good" approximation of the ε-CCS.

In the next subsection, we report an additional motivating example, which further highlights the importance of the problem at hand.

2.2 Motivating example: probabilistic set-membership estimation

Consider the problem of finding $\bar{\theta} \in \mathbb{R}^{n_\theta}$ such that

$$|y - \bar{\theta}^T \varphi(x)| \leq \rho, \ \forall (x, y) \in \mathbb{W} \subseteq \mathbb{R}^{n_x} \times \mathbb{R},$$

where $\varphi : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\theta}$ is a (possibly non-linear) regressor function, and $\rho > 0$ is a given hyperparameter accounting for modelling errors. The (deterministic) set membership estimation problem, see [29–31], consists of computing the set of parameters θ that satisfy the constraint $|y - \theta^T \varphi(x)| \leq \rho$ for all possible values of $(x, y) \in \mathbb{W}$. In the literature, this set is usually referred to as the *feasible parameter set*, that is

$$\mathsf{FPS} \doteq \{ \ \theta \in \mathbb{R}^{n_\theta} \ : \ |y - \theta^T \varphi(x)| \leq \rho, \ \forall (x, y) \in \mathbb{W} \ \}.$$

We notice that FPS could be empty if ρ is chosen too small. If, for given $w = (x, y)$, we define the set

$$\mathbb{X}(w) = \{ \ \theta \in \mathbb{R}^{n_\theta} \ : \ |y - \theta^T \varphi(x)| \leq \rho \ \},$$

then the feasible parameter set FPS can be rewritten as

$$\mathsf{FPS} = \{ \ \theta \in \mathbb{R}^{n_\theta} \ : \ \theta \in \mathbb{X}(w), \ \forall w \in \mathbb{W} \ \}.$$

The deterministic set membership problem suffers from the following limitations in real applications: i) due to the possible non-linearity of $\varphi(\cdot)$, checking if a given $\theta \in \mathbb{R}^{n_\theta}$ satisfies the constraint $\theta \in \mathbb{X}(w)$, for every $w \in \mathbb{W}$, is often a difficult problem; ii) in many situations the robust constraint can not be checked because only samples of $\mathbb{W}$ are available and therefore, only outer bounds of FPS can be computed; and iii) there are problem instances where, because of outliers and possible non-finite support of $\mathbb{W}$, there is no point in $\mathbb{R}^{n_\theta}$ guaranteeing the fulfilment of every possible constraint, and thus, the set FPS is empty (even for large values of ρ).

To deal with this issue, one can resort to a probabilistic relaxation of the FPS. If a probability distribution is defined on $\mathbb{W}$, the probabilistic set membership estimation problem is that of characterizing the set of parameters θ that satisfy

$$\mathsf{Pr}_{\mathbb{W}}\{|y - \theta^T \varphi(x)| \leq \rho\} \geq 1 - \varepsilon,$$

for a given probability parameter $\varepsilon \in (0, 1)$. Hence, we can define FPS_ε as the set of parameters that satisfy the previous probabilistic constraint, that is,

$$\mathsf{FPS}_\varepsilon = \{ \ \theta \in \mathbb{R}^{n_\theta} \ : \ \mathsf{Pr}_{\mathbb{W}}\{\theta \in \mathbb{X}(w)\} \geq 1 - \varepsilon \ \}.$$

It is immediate to notice that this problem fits in the formulation proposed in this section: It suffices to define

$$F(w) = \begin{bmatrix} \varphi^T(x) \\ -\varphi^T(x) \end{bmatrix}, \ g(w) = \begin{bmatrix} \rho + y \\ \rho - y \end{bmatrix}.$$

2.3 Chance-constrained set approximations

Motivated by the discussion above, we are ready to formulate the main problem studied in this paper.

Problem 1 (ε-CCS approximation) *Given the set of linear inequalities* (1)*, and a violation parameter ε, find an inner approximation of the set $\mathbb{X}_\varepsilon$. The approximation should be: i) simple enough; ii) accurate enough, iii) easily computable.*

A solution to this problem is provided in the paper. In particular, regarding i), we present a solution in which the approximating set is represented by *few linear inequalities*. Regarding ii) and iii), we propose a highly tunable and computationally efficient procedure for its construction (see Algorithm 1).

Before presenting our approach, in the next section we provide a brief literature overview of different methods presented in the literature to construct approximations of the ε-CCS set.

3 Overview on different approaches to ε-CCS approximations

The construction of computational efficient approximations to ε-CCS is a long-standing problem. In particular, the reader is referred to the recent reviews [10,32], which provide rather complete discussions on the topic, and cover the most recent results. The authors of [10] distinguish three different approaches, which we very briefly revisit here.

3.1 Exact techniques

In some very special cases, the ε-CCS is convex and hence the CCO problem is efficiently computable. This is the case, for instance, of *individual* chance constraints with w being Gaussian [33]. Other important examples of convexity of the set $\mathbb{X}_\varepsilon$ involve log-concave distribution [1,34]. General sufficient conditions on the convexity of chance constraints may be found in [24, 35–37]. However, all these cases are very specific and hardly extend to the joint chance constraints considered in this work.

All these previously cited references deal with *continuous* distributions. A different line of research concentrates instead on *discrete* distributions, which arise frequently in applications, either directly, or as empirical approximations of the underlying distribution (see, for example, [1, 38]). For this particular case, exact results based on the concept of p-efficiency points [39] or dual methods [40] have been proposed.

3.2 Deterministic approximations

A second class of approaches consist in finding *deterministic* conditions that allow to construct a set $\underline{\mathbb{X}}$, which is a guaranteed inner convex approximation of the probabilistic set $\mathbb{X}_\varepsilon$. The classical solution consists in the applications of Chebyshev-like inequalities, see e.g. [41,42]. More recent techniques, which are proved particularly promising, involve robust optimization [3], as the convex Bernstein-based approximations introduced in [14, 43]. A particular interesting convex relaxation involves the so-called Conditional Value at Risk (CVaR), see [44] and references therein. Finally, we point out some recent techniques based on Genz' code for Gaussian probabilities of rectangles [45], or on polynomial moments relaxations [46, 47].

Specific solutions have been proposed for the case of discrete distributions, see the recent survey [48]. In particular, we point out the recent works proposing a Boolean reformulation of the feasible set of individual and joint chance constraints (see [49, 50]).

Nonetheless, it should be remarked that these techniques usually suffer from conservatism and computational complexity issues, especially in the case of joint chance constraints.

3.3 Sample-based techniques

In recent years, a novel approach to approximate chance constraints, based on random sampling of the uncertain parameters, has gained popularity, see e.g. [4, 5, 51, 52] and references therein. Sampling-based techniques are characterized by the use of a finite number N of iid samples of the uncertainty $\left\{w^{(1)}, w^{(2)}, \ldots, w^{(N)}\right\}$ drawn according to a probability distribution $\mathsf{Pr}_{\mathbb{W}}$. With each sample $w^{(i)}, i \in [N]$, we can associate the following *sampled set*

$$\mathbb{X}(w^{(i)}) = \{\ \theta \in \mathbb{R}^{n_\theta}\ :\ F(w^{(i)})\theta \leq g(w^{(i)})\ \}, \qquad (7)$$

sometimes referred to as *scenario*, since it represents an observed instance of the uncertain constraint.

Then, the scenario approach considers the CCO problem (6) and approximates its solution through the following *scenario problem*

$$\begin{aligned} &\theta^*_{sc} = \arg\min J(\theta) \\ &\text{subject to } \theta \in \mathbb{X}(w^{(i)}), i \in [N]. \end{aligned} \qquad (8)$$

We note that, if the function $J(\theta)$ is convex, problem (8) becomes a linearly constrained convex program, for which very efficient solution approaches exist. Under some technical assumptions (feasibility of the problem and non-degeneracy), a fundamental result [53–56] provides a probabilistic certification of the constraint satisfaction for the solution to the scenario problem. In particular, it is shown that

$$\mathsf{Pr}_{\mathbb{W}^N}\left\{\mathsf{Viol}(\theta^*_{sc}) > \varepsilon\right\} \leq \mathbf{B}(n_\theta - 1; N, \varepsilon), \qquad (9)$$

where the probability in (9) is measured with respect to the samples $\{w^{(1)}, w^{(2)}, \ldots, w^{(N)}\}$.

A few observations are at hand regarding the scenario approach and its relationship with Problem 1. First, if we define the *sampled constraints set* as

$$\mathbb{X}_N \doteq \bigcap_{i=1}^{N} \mathbb{X}(w^{(i)}), \qquad (10)$$

we see that the scenario approach consists in approximating the constraint $\theta \in \mathbb{X}_\varepsilon$ in (6) with its sampled version $\theta \in \mathbb{X}_N$. On the other hand, it should be remarked that the scenario approach cannot be used to derive any guarantee on the relationship existing between $\mathbb{X}_N$ and $\mathbb{X}_\varepsilon$.

Indeed, the nice probabilistic property in (9) holds *only for the optimum of the scenario program* θ^*_{sc}. This is a fundamental point, since the scenario results build on the so-called support constraints, which are defined for the optimum point θ^*_{sc} only.

On the contrary, in our case we are interested in establishing a direct relation (in probabilistic terms) between the set $\mathbb{X}_N$ and the ε-CCS $\mathbb{X}_\varepsilon$. This is indeed possible, but one needs to resort to results based on Statistical Learning Theory [57] and in [52, Theorem 8], summarized in the following lemma.

Lemma 1 (Statistical Learning Theory bound) *Given probabilistic levels $\delta \in (0,1)$ and $\varepsilon \in (0, 0.14)$, if the number of samples N is chosen so that $N \geq N_{LT}$, with*

$$N_{LT} \doteq \frac{4.1}{\varepsilon}\Big(\ln\frac{21.64}{\delta} + 4.39 n_\theta \log_2\Big(\frac{8 e n_\ell}{\varepsilon}\Big)\Big), \quad (11)$$

then $\mathsf{Pr}_{\mathbb{W}^N}\{\mathbb{X}_N \subseteq \mathbb{X}_\varepsilon\} \geq 1-\delta$.

The lemma, whose proof is reported in Appendix A.1, is a direct consequence of the statistical learning theory results on the so-called (α, k)-Boolean functions, given in [52, Corollary 4], where more general results are reported for cases in which ε is not constrained in $(0, 0.14)$.

Remark 2 (Sample-based SMPC) *The learning theory-based approach discussed in this section has been applied in [11] to derive offline a probabilistic inner approximation of the chance-constrained set defining the couples of input/state guaranteeing the desired input/state chance. In particular, the bound* (11) *is a direct extension to the case of joint CC of the result proved in [11] for individual CC. To this regard, we note that the results in the previous section allow to develop a novel SMPC scheme which considers multiple constraints at the same time. These developments are omitted here for brevity, and are reported in an extended version available at [58]. This work highligths the limits of [11]: even for a moderately sized MPC problem with $n_x = 5$ states, $n_u = 2$ inputs, prediction horizon of $T = 10$, simple interval constraints on states and inputs, and for a reasonable choice of probabilistic parameters $\varepsilon = 0.05$, $\delta = 10^{-6}$, we get more than* 1.6 *million linear inequalities.*

Remark 2 motivates the approach presented in the next section, which builds upon the results presented in [15]. We show how the probabilistic scaling approach directly leads to approximations of user-chosen complexity, which can be directly used in applications instead of creating the need for a post-processing step to reduce the complexity of the sampled set.

4 The Probabilistic Scaling Approach

We propose a novel sample-based approach, alternative to the randomized procedures proposed so far. This scheme allows to maintain the nice probabilistic features of these techniques, while at the same time providing the designer with a way of tuning the complexity of the approximation.

The main idea behind this approach consists of a two-step procedure: i) first a simple initial approximation $\theta_c \oplus \mathbb{S}$ of the *shape* of the probabilistic set $\mathbb{X}_\varepsilon$ is obtained, and ii) a *scalable simple approximating set* (Scalable SAS) of the form

$$\mathbb{S}(\gamma) \doteq \theta_c \oplus \gamma\mathbb{S} \quad (12)$$

is considered.

These sets are described by a center point θ_c and a low-complexity shape set $\mathbb{S}$. The center θ_c and the shape $\mathbb{S}$ constitute the *design parameters* of the proposed approach. By appropriately selecting the shape $\mathbb{S}$, the designer can control the complexity of the approximating set. The nonnegative scalar γ controls instead the *scale* of the set $\mathbb{S}(\gamma)$: the larger γ, the larger will be the set.

Note that we do not ask this initial set to have *any* guarantee of probabilistic nature. What we ask is that this set is being able to "capture" somehow the shape of the set $\mathbb{X}_\varepsilon$. Recipes on a possible procedure for constructing this initial set are provided in Section 5. The center θ_c and the set $\mathbb{S}$ constitute the starting point of a scaling procedure, which allows to derive a probabilistic guaranteed approximation of the ε-CCS, as detailed in the next subsection.

4.1 Probabilistic Scaling

In this section, we address the problem of how to scale the set $\mathbb{S}(\gamma)$ around its center θ_c to guarantee, with confidence level $\delta \in (0,1)$, the inclusion of the scaled set into $\mathbb{X}_\varepsilon$. Within this sample-based procedure we assume that N_γ iid samples $\{w^{(1)}, \ldots, w^{(N_\gamma)}\}$ are obtained from $\mathsf{Pr}_{\mathbb{W}}$ and, based on these, we show how to obtain a scalar $\bar\gamma > 0$ such that

$$\mathsf{Pr}_{\mathbb{W}^{N_\gamma}}\{\mathbb{S}(\bar\gamma) \subseteq \mathbb{X}_\varepsilon\} \geq 1-\delta.$$

To this end, we first define the scaling factor associated to a given realization of the uncertainty.

Definition 4 (Scaling factor) *Given a Scalable SAS $\mathbb{S}(\gamma)$, with given center $\theta_c \in \mathbb{R}^{n_\theta}$ and shape $\mathbb{S}$, and a realization $w \in \mathbb{W}$, we define the scaling factor of $\mathbb{S}(\gamma)$ relative to w as*

$$\gamma(w) \doteq \begin{cases} 0 & \text{if } \theta_c \notin \mathbb{X}(w) \\ \max\limits_{\mathbb{S}(\gamma)\subseteq\mathbb{X}(w)} \gamma & \text{otherwise,} \end{cases} \quad (13)$$

with $\mathbb{X}(w) = \{\, \theta \in \mathbb{R}^{n_\theta} \;:\; F(w)\theta \leq g(w) \,\}$.

That is, $\gamma(w)$ represents the maximal scaling that can be applied to $\mathbb{S}(\gamma) = \theta_c \oplus \gamma\mathbb{S}$ around the center θ_c so that $\mathbb{S}(\gamma) \subseteq \mathbb{X}(w)$. The following theorem, whose proof is reported in Appendix A.2, states how to obtain, by means of sampling, a scaling factor $\bar\gamma$ that guarantees, with high probability, that $\mathbb{S}(\bar\gamma) \subseteq \mathbb{X}_\varepsilon$.

Theorem 1 (Probabilistic scaling) *Given a candidate Scalable SAS $\mathbb{S}(\gamma)$, with $\theta_c \in \mathbb{R}^{n_\theta}$ and shape $\mathbb{S}$, suppose that accuracy parameter $\varepsilon \in (0,1)$, confidence level $\delta \in (0,1)$, integer parameter $r \geq 1$ and $N_\gamma \geq r$ are chosen such that*

$$\mathbf{B}(r-1; N_\gamma, \varepsilon) \leq \delta. \tag{14}$$

Draw N_γ iid samples $\{w^{(1)}, w^{(2)}, \ldots, w^{(N_\gamma)}\}$ from distribution $\mathsf{Pr}_{\mathbb{W}}$, compute the corresponding scaling factors

$$\gamma_i \doteq \gamma(w^{(i)}), \quad i \in [N_\gamma]$$

according to Definition 4, and let $\Gamma \doteq \{\gamma_i\}_{i=1}^{N_\gamma}$. Define

$$\bar{\gamma} = \gamma_r^- = \min^{(r)}(\Gamma),$$

i.e. $\bar{\gamma}$ is the r-th smallest value of Γ (see Notation). Under these assumptions:

(i) If $\bar{\gamma} > 0$ then, with probability no smaller than $1-\delta$,

$$\mathbb{S}(\bar{\gamma}) = \theta_c \oplus \bar{\gamma}\mathbb{S} \subseteq \mathbb{X}_\varepsilon.$$

(ii) If $\theta_c \notin \mathbb{X}_\varepsilon$ then $\bar{\gamma} = 0$ with probability no smaller than $1-\delta$.

We now state a Lemma, whose proof is reported in Appendix A.3, which is a generalization of a previous result presented in [19]. The Lemma provides a way to relate the choice of r and N_γ by introducing the value $\beta\epsilon N_\gamma$. This latter represents the desired fraction of the expected number of violations, which can be interpreted as a trade-off parameter between the number of samples and the tightness of the solution.

Lemma 2 *Let $r = \lceil \beta\varepsilon N_\gamma \rceil$, where $\beta \in (0,1)$, and define*

$$\kappa \doteq \left(\frac{\sqrt{\beta} + \sqrt{2-\beta}}{\sqrt{2}(1-\beta)} \right)^2.$$

Then, inequality (14) is satisfied for

$$N_\gamma \geq \frac{\kappa}{\varepsilon} \ln \frac{1}{\delta}. \tag{15}$$

In particular, the choice $\beta = 0.5$ leads to $r = \left\lceil \frac{\varepsilon N_\gamma}{2} \right\rceil$ and $N_\gamma \geq \frac{7.47}{\varepsilon} \ln \frac{1}{\delta}$.

The above result leads to the following simple algorithm, in which we summarise the main steps for constructing the scaled set, and we provide an explicit way of determining parameter r.

Algorithm 1 Probabilistic SAS Scaling

1: Given a candidate Scalable SAS $\mathbb{S}(\gamma)$, and probability levels ε and δ, choose

$$N_\gamma \geq \frac{7.47}{\varepsilon} \ln \frac{1}{\delta} \quad \text{and} \quad r = \left\lceil \frac{\varepsilon N_\gamma}{2} \right\rceil. \tag{16}$$

2: Draw N_γ samples of the uncertainty $w^{(1)}, \ldots, w^{(N_\gamma)}$.
3: **for** $i = 1$ to N_γ **do**
4: Compute, according to Definition 4, the N_γ scaling factors

$$\gamma_i \doteq \gamma(w^{(i)}),\ i \in [N_\gamma]. \tag{17}$$

5: **end for**
6: Return $\bar{\gamma} = \gamma_r^- = \min^{(r)}(\Gamma)$, the r-th smallest value of $\Gamma = \{\gamma_i\}_{i=1}^{N_\gamma}$.

A few comments are in order regarding the algorithm above. In step 4, for each uncertainty sample $w^{(i)}$ one has to solve an optimization problem, which amounts to finding the largest value of γ such that $\mathbb{S}(\gamma)$ is contained in the set $\mathbb{X}(w^{(i)})$ defined in (7). If the SAS is chosen appropriately, we can show that this problem is convex and computationally very efficient: this is discussed in Section 5. Then, in step 6, one has to re-order the sequence $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_{N_\gamma}\}$ so that the first element is the smallest one, the second element is the second smallest one, and so on and so forth, and then return the r-th element of the reordered sequence.

The properties of the output of Algorithm 1 can be derived by a direct application of Theorem 1 and Lemma 2. In particular, if the output $\bar{\gamma}$ is larger than zero, then $\mathbb{S}(\gamma) \subseteq \mathbb{X}_\varepsilon$ with probability no smaller than $1-\delta$.

In the next sections, we provide a "library" of possible candidate SAS shapes. We remind that these sets need to comply with two main requirements: i) being a simple and low-complexity representation; and ii) being able to capture the original shape of the ε-CCS. Moreover, in the light of the discussion after Algorithm 1, we also ask these sets to be convex.

5 Candidate SAS: Sampled-polytope

First, we note that the most straightforward way to design a candidate SAS is again to recur to a sample-based procedure: we draw a fixed number N_S of "design" uncertainty samples [1] $\{\tilde{w}^{(1)}, \ldots, \tilde{w}^{(N_S)}\}$, and construct an initial sampled approximation by introducing

[1] These samples are denoted with a tilde to distinguish them from the samples used in the probabilistic scaling procedure.

the sampled-polytope

$$\mathbb{X}_{N_S} = \bigcap_{j=1}^{N_S} \mathbb{X}(\tilde{w}^{(j)}). \tag{18}$$

Note that the sampled polytope $\mathbb{X}_{N_S}$, by construction, is given by the intersection of $n_\ell N_S$ half-spaces. Hence, we observe that this approach provides very precise control on the final complexity of the approximation, through the choice of the number of samples N_S. However, it is also clear that a choice for which $N_S << N_{LT}$ implies that the probabilistic properties of $\mathbb{X}_{N_S}$ before scaling will probably not meet the required specifications (see Lemma 1). However, we emphasize again that this initial geometry doesn't have nor require any probabilistic guarantees, which are instead provided by the probabilistic scaling discussed in Section 4.1. It should be also remarked that this is only one possible heuristic. For instance, along this line one could as well draw many samples and then apply a clustering algorithm to boil it down to a desired number of samples.

We remark that, in order to apply the scaling procedure, we need to define a *center* around which to apply the scaling procedure. To this end, we could compute the so-called Chebyshev center, which for a given norm $\|\cdot\|_p$, is defined as the center of the largest ball inscribed in $\mathbb{X}_{N_S}$, i.e. $\theta_c = \mathsf{Cheb}_p(\mathbb{X}_{N_S})$. Once the center θ_c has been determined, the scaling procedure can be applied to the *sampled-polytope* SAS defined as

$$\mathbb{S}_{N_S}(\gamma) \doteq \theta_c \oplus \gamma\{\mathbb{X}_{N_S} \ominus \theta_c\}. \tag{19}$$

We note that computing the Chebyshev center of a given polytope is an easy convex optimization problem, for which efficient algorithms exist, see e.g. [18]. A possible alternative would be the analytic center of $\mathbb{X}_{N_S}$, whose computation is even easier (see [18] for further details). Note that the choice of θ_c only affects the goodness of the shape, but we can never know a priori if the analytic center is a better choice than any random center in the candidate SAS.

Example 2 *[Sample-based approximations] To illustrate how the proposed scaling procedure works in practice in the case of sampled-polytope SAS, we revisit Example 1. To this end, a pre-fixed number N_S of uncertainty samples were drawn, and the set of inequalities*

$$F(\tilde{w}^{(j)})\theta \leq g(\tilde{w}^{(j)}), \quad j \in [N_S],$$

with $F(w), g(w)$ defined in (5)*, were constructed, leading to the set $\mathbb{X}_{N_S}$. Then, its Chebyshev center with respect to norm $\|\cdot\|_2$ was computed, and Algorithm 1 was applied to the sampled-polytope SAS $\mathbb{S}_{N_S}(\gamma)$ defined in* (19)*, with $\varepsilon = 0.05$, $\delta = 10^{-6}$, leading to $N_\gamma = 2,065$.*

We note that, in this case, the solution of the optimization problem in (13) *may be obtained by bisection on γ. Indeed, for given γ, checking if $\mathbb{S}_{N_S}(\gamma) \subseteq \mathbb{X}(w^{(i)})$ amounts to solving some simple linear programs.*

Two different situations were considered: a case where the number of inequalities is rather small $N_S = 100$, and a case where the complexity of the SAS is higher, i.e. $N_S = 1,000$. The outcome procedure is illustrated in Figure 2. We can observe that, for a small N_S – Fig. 2(a) – the initial approximation is rather large (although it is contained in $\mathbb{X}_\varepsilon$, we remark that we do not have any guarantee that this will happen). In this case, the probabilistic scaling returns $\gamma = 0.8954$ which is less than one. This means that, in order to obtain a set fulfilling the desired probabilistic guarantees, we need to shrink it around its center. In the second case, for a larger number of sampled inequalities – Fig. 2(b) - the initial set (the red one) is much smaller, and the scaling procedure inflates the set by returning a value of γ greater than one, i.e. $\gamma = 1.2389$. Note that choosing a larger number of samples for the computation of the initial set does not imply that the final set will be a better approximation of the ε-CCS.

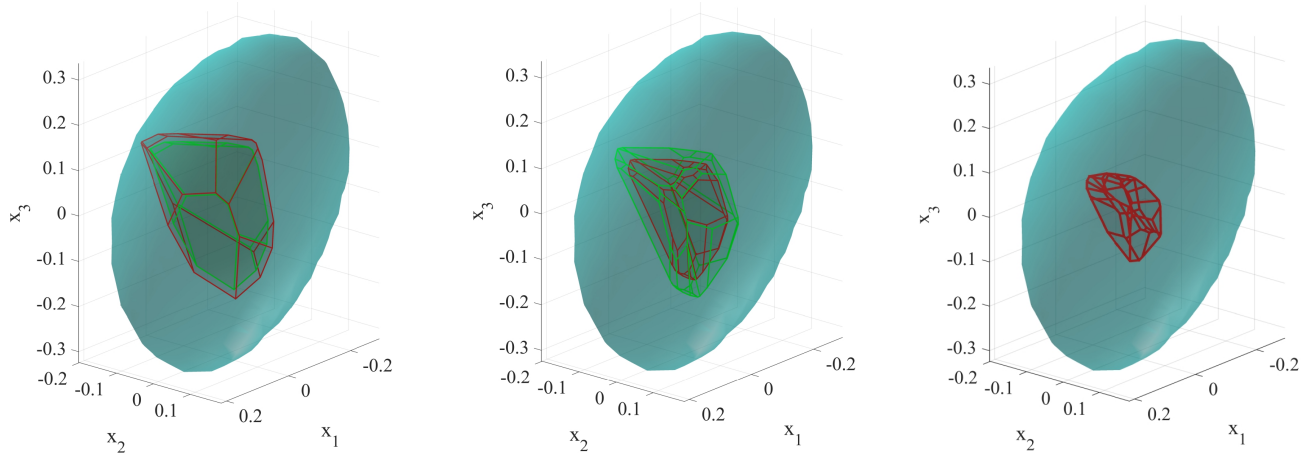
Finally, we compare this approach to the scenario-like ones discussed in Subsection 3.3. To this end, we also draw the approximation obtained by directly applying the Statistical Learning Theory bound (11)*. Note that in this case, since $n_\theta = 3$ and $n_\ell = 4$, we need to take $N_{LT} = 13,011$ samples, corresponding to $52,044$ linear inequalities. The resulting set is represented in Fig. 2(c). We point out that using this approximation i) the set is much more complex, since the number of involved inequalities is much larger; ii) the set is much smaller, hence providing a much more conservative approximation of the ε-CCS. Hence, the ensuing chance-constrained optimization problem will be computationally harder, and lead to a solution with a larger cost or even to an infeasible problem, in cases where the approximating set is too small.*

6 Candidate SAS: Norm-based SAS

In this section, we propose a procedure in which the shape of the scalable SAS is selected a-priori. This corresponds to situations where the designer wants to have full control over the final shape in terms of structure and complexity. The main idea is to define so-called *norm-based SAS* of the form

$$\mathbb{S}_{\ell_p}(\gamma) \doteq \theta_c \oplus \gamma H \mathbb{B}_p^s, \tag{20}$$

where $\mathbb{B}_p^s$ is an ℓ_p-ball in $\mathbb{R}^s$, $H \in \mathbb{R}^{n_\theta \times s}$, with $s \geq n_\theta$, is a design matrix (not necessarily square), and γ is the scaling parameter. Note that when the matrix H is square (i.e. $s = n_\theta$) and positive definite these sets belong to the



(a) $\mathbb{S}_{N_S}$ with $N_S = 100$. $\rightarrow \gamma = 0.8954$ (b) $\mathbb{S}_{N_S}$ with $N_S = 1,000$. $\rightarrow \gamma = 1.2389$ (c) Based on Lemma 1 ($N_{LT} = 52,044$)

Fig. 2. (a-b) Probabilistic scaling approximations of the ε-CCS. Scaling procedure applied to a sampled-polytope with $N_S = 100$ (a) and $N_S = 1,000$ (b). The initial sets are depicted in red, the scaled ones in green. (c) Approximation obtained by direct application of Lemma 1. Note that, in this latter case, to plot the set without out-of-memory errors a pruning procedure [59] of the $52,044$ linear inequalities was necessary.

class of ℓ_p-norm based sets originally introduced in [60]. In particular, in case of ℓ_2 norm, the sets are ellipsoids. This particular choice is the one studied in [16]. Here, we extend this approach to a much more general family of sets, which encompasses for instance *zonotopes*, obtained by letting $p = \infty$ and $s \geq n_\theta$. Zonotopes have been widely studied in geometry, and have found several applications in systems and control, in particular for problems of state estimation and robust Model Predictive Control, see e.g. [61]. Zonotopes proved to be very flexible and yield very efficient implementations.

6.1 Scaling factor computation for norm-based SAS

We recall that the scaling factor $\gamma(w)$ is defined as 0 if $\theta_c \notin \mathbb{X}(w)$ and as the largest value γ for which $\mathbb{S}_{\ell_p}(\gamma) \subseteq \mathbb{X}(w)$ otherwise. The following theorem, whose proof is reported in Appendix A.4, provides a direct and simple way to compute in closed form the scaling factor for a given candidate norm-based SAS.

Theorem 2 (Scaling factor for norm-based SAS) *Given a norm-based SAS $\mathbb{S}(\gamma) = \theta_c \oplus \gamma H \mathbb{B}_p^s$ and a realization $w \in \mathbb{W}$, define $\tau_\ell(w) \doteq g_\ell(w) - f_\ell^T(w)\theta_c$ and $\rho_\ell(w) \doteq \|H^T f_\ell(w)\|_{p^*}$, with $\|\cdot\|_{p^*}$ being the dual norm of $\|\cdot\|_p$.*

The scaling factor $\gamma(w)$ can be computed as

$$\gamma(w) = \min_{\ell \in [n_\ell]} \gamma_\ell(w),$$

with $\gamma_\ell(w)$, $\ell \in [n_\ell]$, given by

$$\gamma_\ell(w) = \begin{cases} 0 & \text{if } \tau_\ell(w) < 0, \\ \infty & \text{if } \tau_\ell(w) \geq 0 \text{ and } \rho_\ell(w) = 0 \\ \dfrac{\tau_\ell(w)}{\rho_\ell(w)} & \text{if } \tau_\ell(w) \geq 0 \text{ and } \rho_\ell(w) > 0 \end{cases}.$$

Note that $\gamma(w)$ is equal to zero if and only if θ_c is not included in the interior of $\mathbb{X}(w)$.

6.2 Construction of a candidate norm-based set

Similarly to Section 5, we first draw a fixed number N_S of "design" uncertainty samples $\{\tilde{w}^{(1)}, \ldots, \tilde{w}^{(N_S)}\}$, and construct an initial sampled approximation (sampled-polytope SAS) $\mathbb{X}_{N_S}$ by means of (18). Again, we consider the Chebyshev center of $\mathbb{X}_{N_S}$, or its analytical center as a possible center θ_c for our approach.

Analogously to what was proposed in [16], given $\mathbb{X}_{N_S}$, $s \geq n_\theta$ and $p \in \{1, 2, \infty\}$, the objective is to compute the largest set $\theta_c \oplus H\mathbb{B}_p^s$ included in $\mathbb{X}_{N_S}$. To this end, we assume that we have a function $\mathsf{Vol}_p(H)$ that provides a measure of the size of $H\mathbb{B}_p^s$. That is, larger values of $\mathsf{Vol}_p(H)$ are obtained for increasing sizes of $H\mathbb{B}_p^s$.

Remark 3 (On the volume function) *The function $\mathsf{Vol}_p(H)$ may be seen as a generalization of the classical concept of Lebesgue volume of the set $\mathbb{X}_{N_S}$. Indeed, when H is a square positive definite matrix, some possibilities are $\mathsf{Vol}_p(H) = \log\det(H)$ – which is directly proportional to the classical volume definition, or $\mathsf{Vol}_p(H) = \mathrm{tr}\,\mathrm{H}$ – which for $p = 2$ becomes the well known*

sum of ellipsoid semiaxes (see [62] and [18, Chapter 8]). These measures can be easily generalized to non square matrices. It suffices to compute the singular value decomposition. If $H = U\Sigma V^T$, we could use the measures $\mathsf{Vol}_p(H) = \operatorname{tr}\Sigma$ or $\mathsf{Vol}_p(H) = \log\det(\Sigma)$.
For non square matrices H, specific results for particular values of p are known. For example, we remind that if $p = \infty$ and $H \in \mathbb{R}^{n_\theta \times s}$, $s \geq n_\theta$, then $\theta_c \oplus H\mathbb{B}^s_\infty$ is a zonotope. Then, if we denote as generator each of the columns of H, the volume of a zonotope can be computed by means of a sum of terms (one for each different way of selecting n_θ generators out of the s generators of H); see [63], [64]. Another possible measure of the size of a zonotope $\theta_c \oplus H\mathbb{B}^s_\infty$ is the Frobenious norm of H [63].

Given an initial design set $\mathbb{X}_{N_S}$, we elect as our candidate Scalable SAS the largest "volume" norm-based SAS contained in $\mathbb{X}_{N_S}$. Formally, this rewrites as the following optimization problem

$$\begin{aligned} &\max_{\theta_c, H} \ \mathsf{Vol}_p(H) \\ &\text{subject to } \theta_c \oplus H\mathbb{B}^s_p \subseteq \mathbb{X}_{N_S}. \end{aligned} \tag{21}$$

As it has been shown (see Appendix A.4), problem (21) is equivalent to

$$\begin{aligned} &\min_{\theta_c, H} \ -\mathsf{Vol}_p(H) \\ &\text{s.t. } f_\ell^T(\tilde{w}^{(j)})\theta_c + \|H^T f_\ell(\tilde{w}^{(j)})\|_{p^*} - g_\ell(\tilde{w}^{(j)}) \leq 0, \\ &\qquad \ell \in [n_\ell],\ j \in [N_S], \end{aligned} \tag{22}$$

where we have replaced the maximization of $\mathsf{Vol}_p(H)$ with the minimization of -$\mathsf{Vol}_p(H)$.

We notice that the constraints are convex on the decision variables; also, the functional to minimize is convex under particular assumptions. For example when H is assumed to be square and positive definite and $\mathsf{Vol}_p(H) = \log\det(H)$. For non square matrices, the constraints remain convex, but the convexity of the functional to be minimized is often lost. In this case, local optimization algorithms should be employed to obtain a possibly sub-optimal solution.

Example 3 (Norm-based SAS) *We revisit again Example 1 to show the use of norm-based SAS. We note that, in this case, the designer can control the approximation outcome by acting upon the number of design samples N_S used for constructing the set $\mathbb{S}_{N_S}$. In Figure 3 we report two different norm-based SAS, respectively with $p = 1$ and $p = \infty$, and for each of them we consider two different values of N_S, respectively $N_S = 100$ and $N_S = 1,000$. Similarly to what observed in Example 2, we see that for larger N_S, the ensuing initial set becomes smaller. Consequently, we have a shrinkage process for small N_S and an inflating one for large N_S. However, we observe that in this case, the final number of inequalities is independent of N_S (8 for $\mathbb{S}_{\ell_1}$ and 6 for $\mathbb{S}_{\ell_\infty}$).*

6.2.1 Relaxed computation

It is worth remarking that the minimization problem of the previous subsection might be infeasible. In order to guarantee the feasibility of the problem, a soft-constrained optimization problem is proposed. With a relaxed formulation, θ_c is not guaranteed to satisfy all the sampled constraints. However $\theta_c \in \mathbb{S}_{N_S}$ is not necessary to obtain an ε-CCS.

Given $\xi > 0$, the relaxed version of optimization problem (22) is

$$\begin{aligned} &\min_{\theta_c, H, \eta_1, \ldots, \eta_{N_S}} \ -\mathsf{Vol}_p(H) + \xi \sum_{j=1}^{N_S} \max\{\eta_j, 0\} \\ &\text{s.t. } f_\ell^T(\tilde{w}^{(j)})\theta_c + \|H^T f_\ell(\tilde{w}^{(j)})\|_{p^*} - g_\ell(\tilde{w}^{(j)}) \leq \eta_j, \\ &\qquad \ell \in [n_\ell],\ j \in [N_S]. \end{aligned} \tag{23}$$

The parameter ξ serves to provide an appropriate trade off between satisfaction of the sampled constraints and the size of the obtained region. A possibility to choose ξ would be to choose it in such a way that the fraction of violations n_{viol}/N_S (where n_{viol} is the number of elements η_j larger than zero) is smaller than $\varepsilon/2$.

7 Numerical example: Probabilistic set membership estimation

We now present a numerical example in which the results of the paper are applied to the probabilistic set membership estimation problem, introduced in subsection 2.2. We consider the universal approximation functions given by Gaussian radial basis function networks (RBFN) [65].

Given the nodes $[x_1, x_2, \ldots, x_M]$ and the variance parameter c, the corresponding Gaussian radial basis function network is defined as

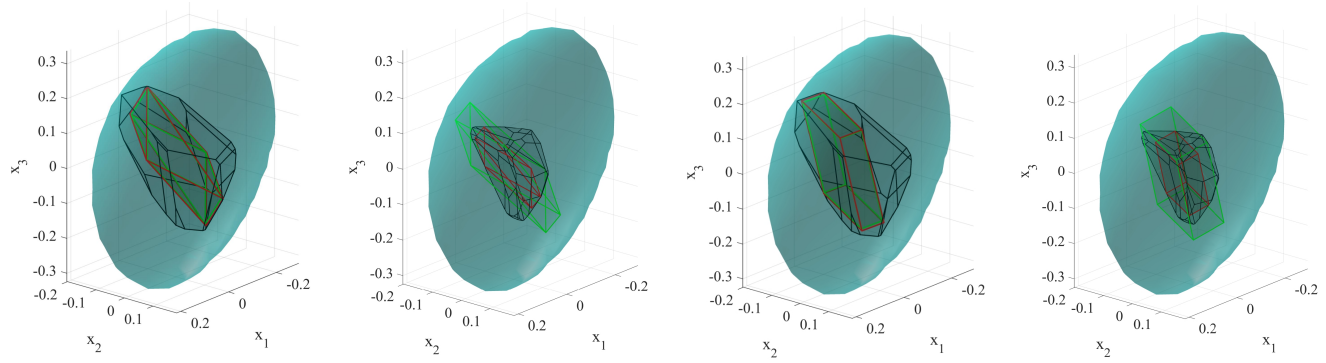
$$\mathrm{RBFN}(x, \theta) = \theta^T \varphi(x),$$

where $\theta = \begin{bmatrix} \theta_1 & \ldots & \theta_M \end{bmatrix}^T$ represents the weights and

$$\varphi(x) = \begin{bmatrix} \exp\left(\frac{-\|x - x_1\|^2}{c}\right) & \ldots & \exp\left(\frac{-\|x - x_M\|^2}{c}\right) \end{bmatrix}^T$$

is the regressor function. Given $\delta \in (0,1)$ and $\varepsilon \in (0,1)$, the objective is to obtain, with probability no smaller than $1 - \delta$, an inner approximation of the probabilistic feasible parameter set FPS_ε, which is the set of parameters $\theta \in \mathbb{R}^M$ that satisfies

$$\mathsf{Pr}_{\mathbb{W}}\{|y - \theta^T \varphi(x)| \leq \rho\} \geq 1 - \varepsilon, \tag{24}$$



(a) $\gamma = 0.9701$ (b) $\gamma = 1.5995$ (c) $\gamma = 0.9696$ (d) $\gamma = 1.5736$

Fig. 3. Scaling procedure applied to (a) $\mathbb{S}_{\ell_1}$-SAS with $N_S = 100$, (b) $\mathbb{S}_{\ell_1}$-SAS with $N_S = 1,000$, (c) $\mathbb{S}_{\ell_\infty}$-SAS with $N_S = 100$, and (d) $\mathbb{S}_{\ell_\infty}$-SAS with $N_S = 1,000$. The initial set is depicted in red, the final one in green. The sampled design polytope $\mathbb{S}_{N_S}$ is represented in black.

where $\rho = 5$, x is a random scalar with uniform distribution in $[-5, 5]$, and

$$y = \sin(3x) + \sigma,$$

where σ is a random scalar with a normal distribution with mean 5 and variance 1.

We use the procedure detailed in Sections 4, 5 and 6 to obtain a SAS of FPS_ε. We have taken a grid of $M = 20$ points in the interval $[-5, 5]$ to serve as nodes for the RBFN, and a variance parameter of $c = 0.15$. We have taken $N_S = 350$ random samples $w = (x, y)$ to compute the initial geometry, which has been chosen to be an ℓ_∞ norm-based SAS of dimension 20 with a relaxation parameter of $\xi = 1$ (see (23)). The chosen initial geometry is $\theta_c \oplus H\mathbb{B}_\infty^{20}$, where H is constrained to be a diagonal matrix.

When the initial geometry is obtained, we scale it around its center by means of probabilistic scaling with Algorithm 1. The number of samples required for the scaling phase to achieve $\varepsilon = 0.05$ and $\delta = 10^{-6}$ is $N_\gamma = 2065$ and the resulting scaling factor is $\gamma = 0.3803$. The scaled geometry $\theta_c \oplus \gamma H\mathbb{B}_\infty^{20}$ is, with a probability no smaller than $1-\delta$, an inner approximation of FPS_ε which we will refer to as $\mathsf{FPS}_\varepsilon^\delta$. Since it is a transformation of an ℓ_∞ norm ball with a diagonal matrix H, we can write it as

$$\mathsf{FPS}_\varepsilon^\delta = \{\theta : \theta^- \leq \theta \leq \theta^+\},$$

where the extreme values $\theta^-, \theta^+ \in \mathbb{R}^{20}$ are represented in Figure 4, along with the central value $\theta_c \in \mathbb{R}^{20}$.

Once the $\mathsf{FPS}_\varepsilon^\delta$ has been computed, we can use its center θ_c to make the point estimation $y \approx \theta_c^T \varphi(x)$. We can also obtain probabilistic upper and lower bounds of y by means of equation (24). That is, every point in $\mathsf{FPS}_\varepsilon^\delta$

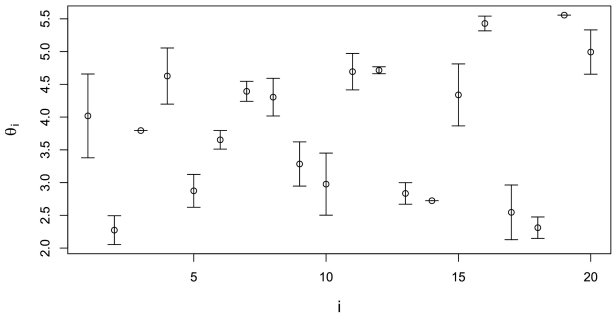


Fig. 4. Representation of the extreme values θ^+ and θ^- and the central value θ_c of the $\mathsf{FPS}_\varepsilon^\delta$.

satisfies, with confidence $1 - \delta$:

$$\begin{aligned} \mathsf{Pr}_{\mathbb{W}}\{y \leq \theta^T \varphi(x) + \rho\} \geq 1 - \varepsilon, \\ \mathsf{Pr}_{\mathbb{W}}\{y \geq \theta^T \varphi(x) - \rho\} \geq 1 - \varepsilon. \end{aligned} \tag{25}$$

We notice that the tightest probabilistic bounds are obtained with θ^+ for the lower bound and θ^- for the upper one. That is, we finally obtain that, with confidence $1-\delta$:

$$\begin{aligned} \mathsf{Pr}_{\mathbb{W}}\{y \leq {\theta^-}^T \varphi(x) + \rho\} \geq 1 - \varepsilon, \\ \mathsf{Pr}_{\mathbb{W}}\{y \geq {\theta^+}^T \varphi(x) - \rho\} \geq 1 - \varepsilon. \end{aligned} \tag{26}$$

Figure 5 shows the results of both the point estimation and the probabilistic interval estimation.

8 Concluding remarks

In this paper, we proposed a general approach to construct probabilistically guaranteed inner approximations of the chance-constrained set $\mathbb{X}_\varepsilon$. The approach is very general and flexible. In this section, we report

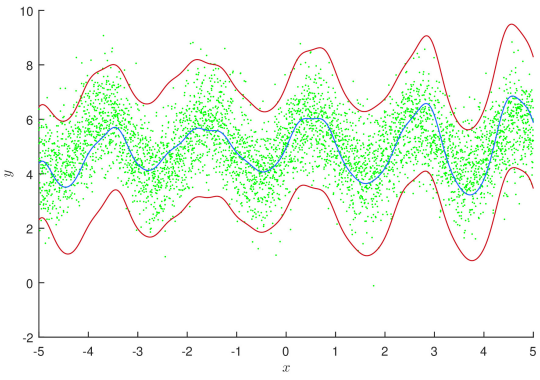


Fig. 5. Real values of y vs central estimation (blue) and interval prediction bounds (red).

a few final remarks on some important aspects of the presented methodology.

8.1 On scalability of the proposed approach

We point out that our framework provides different schemes, with different computational requirements. In particular, regarding the norm-based sets discussed in Section 6, Theorem 2 provides a closed-form expression for the scaling computations. Hence, the approach scales extremely well when the initial candidate set $\theta_c \oplus H\mathbb{B}_p^s$ is given. In the case the initial set is for instance a generic polytope (as for the sampled-polytopes discussed in Section 5), the scaling computation is indeed more involved, since there is not a close-form expression in general. In this case, the solution of the optimization problem in (17) may be obtained by bisection on γ. Note that, in this case, for given γ, checking if $\mathbb{S}_{N_S}(\gamma) \subseteq \mathbb{X}(w^{(i)})$ amounts to solving a linear program.

Otherwise, when the set $\theta_c \oplus H\mathbb{B}_p^s$ is not available, its computation will clearly constitute the most demanding step of our scheme. In this case, as detailed in Section 6, θ_c and H can be obtained by means of a convex optimization problem when H is a square matrix. Depending on the choice on H, the number of decision variables increases linearly with the dimension of θ_c (e.g. H is a diagonal matrix), or quadratically (if H is a full matrix). Again, we stress that the richer is the family of initial candidate sets (e.g. when the initial set is a zonotope), the more demanding will be its computation.

In any case, we are not claiming that the approach we propose is to be preferred to other approaches in every situation. For instance, if the uncertainty enters in a "nice" way, better solutions surely exist. And of course, there will be situations where the solutions discussed in Section 3 may be preferable. On the other hand, a nice and distinctive feature of our approach is that it can be seen as *complementary* to these approaches: for instance, imagine one has constructed a *safe tractable approximation* based on the procedure proposed in [43], i.e. an approximation given by a convex set [43, Proposition 1] of the form $\mathcal{H}(\theta) \leq 0$. Then, nothing forbids the designer to use this set as initial SAS to which applying our scaling procedure, thus further improving its approximation properties.

Moreover, it should be remarked that the tunability of our method, while allowing high flexibility, entails by definition the problem of parameter selection. In our case, the main degree of freedom is the choice of the initial *scalable set*. In this case, the trade-off is evident: the more complex the set, the better may be the obtained approximation, at the expense of a possibly larger computational effort. Besides this clear implication, a more detailed analysis, both theoretical and experimental, is needed to understand the effect of specific choices of the initial set (as those introduced in Section 5 and 6). This is an important point that however goes beyond the scope of the present paper, and is the subject of ongoing research.

8.2 Extensions to nonlinear setups

We remark that the proposed scaling approach is not limited to sets defined by linear inequalities, but may be extended to more general sets using very similar arguments. Indeed, we may consider a generic binary performance function $\psi : \mathbb{R}^{n_\theta} \times \mathbb{W} \to \{0, 1\}$ defined as [2]

$$\psi(\theta, w) = \begin{cases} 0 \text{ if } \theta \text{ meets design specifications for } w \\ 1 \text{ otherwise.} \end{cases} \tag{27}$$

In this case, the violation probability may be written as $\mathsf{Viol}(\theta) \doteq \mathsf{Pr}_{\mathbb{W}}\{\psi(\theta, w) = 1\}$, and we can still define the set $\mathbb{X}_\varepsilon$ as in (3). Then, given an initial SAS candidate, Algorithm 1 still provides a valid approximation. However, it should be remarked that, even if we choose a "nice" SAS as those previously introduced, the nonconvexity of ψ will most probably render step 4 of the algorithm intractable. ~~Also, sample-size bounds different from (16) would need to be derived, due to nonconvexity.~~ To further elaborate on this point, let us focus on the case when the design specification may be expressed as a (nonlinear) inequality of the form

$$\zeta(\theta, w) \leq 0.$$

Then, the computation of each scaling factor γ_i of step 4 consists, provided that $\theta_c \in \mathbb{X}(w^{(i)})$, in solving the fol-

[2] Clearly, this formulation encompasses the setup discussed, obtained by simply setting $\psi(\theta, w) = \begin{cases} 0 \text{ if } F(w)\theta \leq g(w) \\ 1 \text{ otherwise.} \end{cases}$

lowing nonconvex optimization problem

$$\gamma_i \doteq \arg\max \gamma$$
$$\text{s.t. } \theta_c \oplus \gamma\mathbb{S} \subseteq \mathbb{X}(w^{(i)}) = \Big\{\theta \in \mathbb{R}^{n_\theta} \mid \zeta(\theta, w^{(i)}) \leq 0\Big\}.$$

We note that this is generally a hard problem. However, there are cases when this problem is still solvable. In particular, we remark that whenever $\zeta(\theta, w)$ is a *convex function* of θ for fixed w and the set $\mathbb{S}$ is also convex, the above optimization problem may be formulated as a convex program. ~~Moreover, in this case, the bound (16) would still hold. Hence, in such situations, the approach proposed here is still completely viable, since all the derivations continue to hold.~~

8.3 Future directions

In the previous subsection, we discussed how the method might be extended to nonlinear setups. These extensions always consider continuous variables. One may wonder if the approach may also be extended to the important class of problems involving *integer* values, such as the mixed-integer programs studied in [38]. This is a problem currently under investigation, however we remark that the extension in this case is far from being trivial, and while we understand that the presented approach is generalizable in theory, we have not yet found any computationally efficient implementation for it.

We also remark that the paper opens the way to the design of other families of Scalable SAS. For instance, we are currently working on using the family of sets defined in the form of polynomial superlevel sets (PSS) proposed in [66].

A Appendix

A.1 Proof of Lemma 1

To prove the lemma, we first recall the following definition from [52].

Definition 5 ((α, k)-Boolean function) *The function $h : \mathbb{R}^{n_\theta} \times \mathbb{W} \to \mathbb{R}$ is an (α, k)-Boolean function if for fixed w it can be written as an expression consisting of Boolean operators involving k polynomials $p_1(\theta), p_2(\theta), \ldots, p_k(\theta)$, in the components θ_i, $i \in [n_\theta]$ and the degree with respect to θ_i of all these polynomials is no larger than α.*

Let us now define the binary functions

$$h_\ell(\theta, w) \doteq \begin{cases} 0 & \text{if } f_\ell(w)\theta \leq g_\ell(w) \\ 1 & \text{otherwise} \end{cases}, \ \ell \in [n_\ell].$$

Introducing the function $h(\theta, w) \doteq \max\limits_{\ell=1,\ldots,n_\ell} h_\ell(\theta, w)$, we see that the violation probability can be alternatively written as $\mathsf{Viol}(\theta) \doteq \mathsf{Pr}_{\mathbb{W}}\{\, h(\theta, w) = 1 \,\}$. We notice that $h(\theta, w)$ is an $(1, n_\ell)$-Boolean function, since it can be expressed as a function of n_ℓ Boolean functions, each of them involving a polynomial of degree 1. The proof now follows from Theorem 8 in [52] that states that if $h : \mathbb{R}^{n_\theta} \times \mathbb{W} \to \mathbb{R}$ is an (α, k)-Boolean function and $\varepsilon \in (0, 0.14)$ then, with probability greater than $1 - \delta$, we have $\mathsf{Pr}_{\mathbb{W}}\{\, h(\theta, w) = 1 \,\} \leq \varepsilon$ if N is chosen such that

$$N \quad \geq \quad \frac{4.1}{\varepsilon}\Big(\ln\frac{21.64}{\delta} \ + \ 4.39 n_\theta \log_2\Big(\frac{8e\alpha k}{\varepsilon}\Big)\Big).$$

■

A.2 Proof of Theorem 1

To prove the theorem, we first prove the following property.

Property 1 *Given $\varepsilon \in (0, 1)$, $\delta \in (0, 1)$, and $r \geq 1$, let $N \geq r$ be such that $\mathbf{B}(r - 1; N, \varepsilon) \leq \delta$. Draw N iid samples $\{w^{(1)}, w^{(2)}, \ldots, w^{(N)}\}$ from a distribution $\mathsf{Pr}_{\mathbb{W}}$. For $i \in [N]$, let $\gamma_i \doteq \gamma(w^{(i)})$, with $\gamma(\cdot)$ as in Definition 4, and suppose that $\bar{\gamma} = \min^{(r)}\{\gamma_i\}_{i=1}^N > 0$. Then, with probability no smaller than $1 - \delta$, it holds that $\mathsf{Pr}_{\mathbb{W}}\{\theta_c \oplus \bar{\gamma}\mathbb{S} \not\subseteq \mathbb{X}(w)\} \leq \varepsilon$.*

Proof: It has been proven in [55, 56] that if one discards no more than s constraints on a convex problem with N random constraints, then the probability of violating the constraints with the solution obtained from the random convex problem is no larger than $\varepsilon \in (0, 1)$, with probability no smaller than $1 - \delta$, where

$$\delta = \binom{s + d - 1}{s} \sum_{i=0}^{s+d-1} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i},$$

and d is the number of decision variables. We apply this result to the following optimization problem

$$\max_{\gamma} \gamma \text{ subject to } \theta_c \oplus \gamma\mathbb{S} \subseteq \mathbb{X}(w^{(i)}), \ \ i \in [N]. \quad \text{(A.1)}$$

From Definition 4, we could rewrite this optimization problem as

$$\max_{\gamma} \gamma \text{ subject to } \gamma \leq \gamma(w^{(i)}), \ i \in [N].$$

We first notice that the problem under consideration is convex and has a unique scalar decision variable γ. That is, $d = 1$. Also, the non-degeneracy and uniqueness assumption required in the application of the results of [55]

and [56] are satisfied. We notice that $\bar{\gamma} = \min^{(r)}\{\gamma_i\}_{i=1}^N$, is the optimal solution to the optimization problem when $s = r-1$ constraints are discarded. Thus, we have that with probability no smaller than $1-\delta$, where

$$\delta = \binom{r-1}{r-1} \sum_{i=0}^{r-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} = \mathbf{B}(r-1; N, \varepsilon),$$

the choice $\bar{\gamma} = \min^{(r)}\{\gamma_i\}_{i=1}N$ satisfies $\mathsf{Pr}_{\mathbb{W}}\{\bar{\gamma} > \gamma(w)\} \leq \varepsilon$.

We conclude from this, and Definition 4, that with probability no smaller than $1-\delta$, $\mathsf{Pr}_{\mathbb{W}}\{\theta_c \oplus \bar{\gamma}\mathbb{S} \not\subseteq \mathbb{X}(w)\} \leq \varepsilon$. ■

Proof of Theorem 1

We consider first the case $\bar{\gamma} > 0$. From Property 1, we have that $\bar{\gamma} > 0$ satisfies, with probability no smaller than $1-\delta$, that $\mathsf{Pr}_{\mathbb{W}}\{\mathbb{S}(\bar{\gamma}) \not\subseteq \mathbb{X}(w)\} \leq \varepsilon$. Equivalently, $\mathsf{Pr}_{\mathbb{W}}\{\mathbb{S}(\bar{\gamma}) \subseteq \mathbb{X}(w)\} \geq 1-\varepsilon$. This can be rewritten as $\mathsf{Pr}_{\mathbb{W}}\{F(w)\theta \leq g(w), \ \forall \theta \in \mathbb{S}(\gamma)\} \geq 1-\varepsilon$, and it implies that the probability of violation in $\theta_c \oplus \bar{\gamma}\mathbb{S}$ is no larger than ε, with probability no smaller than $1-\delta$. This proves the first claim.
Suppose now that $\theta_c \notin \mathbb{X}_\varepsilon$. This is equivalent to $\mathsf{Viol}(\theta_c) = \bar{\varepsilon}_c > \varepsilon$. Suppose that the sample constraints $\theta_c \in \mathbb{X}(w^{(i)}), \ i \in [N_\gamma]$ are violated v_c times. This would imply, because of the definition of scaling factor, that there are at least v_c scaling factors $\gamma(w^{(i)})$ equal to zero. From this and $\mathsf{Viol}(\theta_c) = \bar{\varepsilon}_c > \varepsilon$, we obtain

$$\begin{aligned}\mathsf{Pr}_{\mathbb{W}^{N_\gamma}}\{\bar{\gamma} > 0\} &= \mathsf{Pr}_{\mathbb{W}^{N_\gamma}}\{\min^{(r)}\{\gamma_i\}_{i=1}^{N_\gamma} > 0\} \\ &\leq \mathsf{Pr}_{\mathbb{W}^{N_\gamma}}\{v_c < r\} \\ &= B(r-1; N_\gamma, \varepsilon_c) \leq B(r-1; N_\gamma, \varepsilon) \leq \delta.\end{aligned}$$

From here we conclude that $\theta_c \notin \mathbb{X}_\varepsilon$ implies

$$\mathsf{Pr}_{\mathbb{W}^{N_\gamma}}\{\bar{\gamma} = 0\} = 1 - \mathsf{Pr}_{\mathbb{W}^{N_\gamma}}\{\bar{\gamma} > 0\} \geq 1-\delta. \quad ■$$

A.3 Proof of Lemma 2

From [67, Corollary 1] we have that (14) is satisfied for

$$N_\gamma \geq \frac{1}{\varepsilon}\left(r - 1 + \ln\frac{1}{\delta} + \sqrt{2(r-1)\ln\frac{1}{\delta}}\right). \qquad \text{(A.2)}$$

Since $r - 1 = \lceil \beta\varepsilon N_\gamma \rceil - 1 \leq \beta\varepsilon N_\gamma$, we obtain the sufficient condition

$$\begin{aligned}N_\gamma \geq & \frac{1}{\varepsilon}\left(\beta\varepsilon N_\gamma + \ln\frac{1}{\delta} + \sqrt{2\beta\varepsilon N_\gamma \ln\frac{1}{\delta}}\right) \\ = \ & \beta N_\gamma + \frac{1}{\varepsilon}\ln\frac{1}{\delta} + \sqrt{2\beta N_\gamma \frac{1}{\varepsilon}\ln\frac{1}{\delta}}.\end{aligned}$$

Letting $a \doteq \sqrt{N_\gamma}$ and $b \doteq \sqrt{\frac{1}{\varepsilon}\ln\frac{1}{\delta}}$, the previous expression can be rewritten as $(1-\beta)a^2 - (\sqrt{2\beta}b)a - b^2 \geq 0$. The largest root of this second order equation is

$$\left(\frac{\sqrt{\beta} + \sqrt{2-\beta}}{\sqrt{2}(1-\beta)}\right) b.$$

Thus, (A.2) is satisfied if

$$\sqrt{N_f} \geq \left(\frac{\sqrt{\beta} + \sqrt{2-\beta}}{\sqrt{2}(1-\beta)}\right)\sqrt{\frac{1}{\varepsilon}\ln\frac{1}{\delta}}.$$

This proves the claim.

■

A.4 Proof of Theorem 2

Note that, by definition, the condition $\theta_c \oplus \gamma H\mathbb{B}_p^s \subseteq \mathbb{X}(w)$ is equivalent to

$$\max_{z \in \mathbb{B}_p^s} f_\ell^T(w)(\theta_c + \gamma H z) - g_\ell(w) \leq 0, \ \ell \in [n_\ell].$$

Equivalently, from the dual norm definition, we have

$$f_\ell^T(w)\theta_c + \gamma\|H^T f_\ell(w)\|_{p^*} - g_\ell(w) \leq 0, \ \ell \in [n_\ell].$$

Denote by γ_ℓ the scaling factor γ_ℓ corresponding to the ℓ-th constraint

$$f_\ell^T(w)\theta_c + \gamma_\ell\|H^T f_\ell(w)\|_{p^*} - g_\ell(w) \leq 0.$$

With the notation introduced in the theorem, this constraint rewrites as $\gamma_\ell \rho_\ell(w) \leq \tau_\ell(w)$. The result follows noting that the corresponding scaling factor $\gamma_\ell(w)$ can be computed as

$$\gamma_\ell(w) = \max_{\gamma_\ell \rho_\ell(w) \geq \tau_\ell(w)} \gamma_\ell,$$

and that the value for $\gamma(w)$ is obtained from the most restrictive one. ■

References

[1] A. Prékopa, Stochastic Programming. Kluwer Academic Publishers, 1995.

[2] N. V. Sahinidis, "Optimization under uncertainty: state-of-the-art and opportunities," Computers & Chemical Engineering, vol. 28, no. 6-7, pp. 971–983, 2004.

[3] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," Mathematics of Operations Research, vol. 23, pp. 769–805, 1998.

[4] G. Calafiore, F. Dabbene, and R. Tempo, "Research on probabilistic methods for control system design," Automatica, vol. 47, pp. 1279–1293, 2011.

[5] R. Tempo, G. Calafiore, and F. Dabbene, Randomized Algorithms for Analysis and Control of Uncertain Systems: with Applications. Springer Science & Business Media, 2012.

[6] M. Mammarella, E. Capello, F. Dabbene, and G. Guglieri, "Sample-based SMPC for tracking control of fixed-wing UAV," IEEE Control Systems Letters, vol. 2, no. 4, pp. 611–616, 2018.

[7] J. Li, W. Zhan, Y. Hu, and M. Tomizuka, "Generic tracking and probabilistic prediction framework and its application in autonomous driving," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 9, pp. 3634–3649, 2020.

[8] M. Chamanbaz, F. Dabbene, and C. Lagoa, Algorithms for Optimal AC Power Flow in the Presence of Renewable Sources. Wiley Encyclopedia of Electrical and Electronics Engineering, 2020, pp. 1–13.

[9] M. Chamanbaz, F. Dabbene, and C. M. Lagoa, "Probabilistically robust AC optimal power flow," IEEE Transactions on Control of Network Systems, vol. 6, no. 3, pp. 1135–1147, 2019.

[10] X. Geng and L. Xie, "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization," Annual Reviews in Control, vol. 47, pp. 341–363, 2019.

[11] M. Lorenzen, F. Dabbene, R. Tempo, and F. Allgöwer, "Stochastic MPC with offline uncertainty sampling," Automatica, vol. 81, no. 1, pp. 176–183, 2017.

[12] M. Mammarella, M. Lorenzen, E. Capello, H. Park, F. Dabbene, G. Guglieri, M. Romano, and F. Allgöwer, "An offline-sampling SMPC framework with application to autonomous space maneuvers," IEEE Transactions on Control Systems Technology, vol. 28, no. 2, pp. 1–15, 2018.

[13] 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.

[14] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," SIAM Journal on Optimization, vol. 17, no. 4, pp. 969–996, 2006.

[15] T. Alamo, V. Mirasierra, F. Dabbene, and M. Lorenzen, "Safe approximations of chance constrained sets by probabilistic scaling," in 2019 18th European Control Conference (ECC). IEEE, 2019, pp. 1380–1385.

[16] M. Mammarella, T. Alamo, F. Dabbene, and M. Lorenzen, "Computationally efficient stochastic MPC: a probabilistic scaling approach," in Proc. of 4th IEEE Conference on Control Technology and Applications (CCTA), 2020, pp. 25–30.

[17] T. Alamo, J. Manzano, and E. Camacho, "Robust design through probabilistic maximization," in Uncertainty in Complex Networked Systems. In Honor of Roberto Tempo, T. Basar, Ed. Birkhäuser, 2018, pp. 247–274.

[18] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.

[19] V. Mirasierra, M. Mammarella, F. Dabbene, and T. Alamo, "Prediction error quantification through probabilistic scaling," IEEE Control Systems Letters, vol. 6, pp. 1118–1123, 2022.

[20] L. Hewing, A. Carron, K. P. Wabersich, and M. N. Zeilinger, "On a correspondence between probabilistic and robust invariant sets for linear systems," in 2018 European Control Conference (ECC). IEEE, 2018, pp. 1642–1647.

[21] B. Miller and H. Wagner, "Chance constrained programming with joint constraints," Operations Research, vol. 13, pp. 930–945, 1965.

[22] L. Khachiyan, "The problem of calculating the volume of a polyhedron is enumerably hard," Russian Mathematical Surveys, vol. 44, pp. 199–200, 1989.

[23] A. Shapiro, D. Dentcheva, and A. Ruszczyński, Lectures on stochastic programming: modeling and theory. SIAM, 2014.

[24] W. van Ackooij, "Eventual convexity of chance constrained feasible sets," Optimization, vol. 64, no. 5, pp. 1263–1284, 2015.

[25] W. Van Ackooij and J. Malick, "Eventual convexity of probability constraints with elliptical distributions," Mathematical Programming, vol. 175, no. 1, pp. 1–27, 2019.

[26] A. Prékopa, T. Rapcsák, and I. Zsuffa, "Serially linked reservoir system design using stochastic programing," Water Resources Research, vol. 14, no. 4, pp. 672–678, 1978.

[27] D. Dentcheva, B. Lai, and A. Ruszczyński, "Dual methods for probabilistic optimization problems," Mathematical Methods of Operations Research, vol. 60, no. 2, pp. 331–346, 2004.

[28] M. Lorenzen, F. Dabbene, R. Tempo, and F. Allgöwer, "Constraint-tightening and stability in stochastic model predictive control," IEEE Transactions on Automatic Control, vol. 62, no. 7, pp. 3165–3177, 2017.

[29] A. Vicino and G. Zappa, "Sequential approximation of feasible parameter sets for identification with set membership uncertainty," IEEE Transactions on Automatic Control, vol. 41, no. 6, pp. 774–785, 1996.

[30] J. M. Bravo, T. Alamo, and E. F. Camacho, "Bounded error identification of systems with time-varying parameters," IEEE Transactions on Automatic Control, vol. 51, no. 7, pp. 1144–1150, 2006.

[31] V. Puig, "Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies," International Journal of Applied Mathematics and Computer Science, vol. 20, no. 4, pp. 619–635, 2010.

[32] M. A. Lejeune and A. Prékopa, "Relaxations for probabilistically constrained stochastic programming problems: review and extensions," Annals of Operations Research, pp. 1–22, 2018.

[33] S. Kataoka, "A stochastic programming model," Econometrica: Journal of the Econometric Society, vol. 31, no. 1-2, pp. 181–196, 1963.

[34] A. Prékopa, "Logarithmic concave measures with application to stochastic programming," Acta Scientiarum Mathematicarum, vol. 32, pp. 301–316, 1971.

[35] C. M. Lagoa, "On the convexity of probabilistically constrained linear programs," in Proceedings of the 38th IEEE Conference on Decision and Control, vol. 1, 1999, pp. 516–521.

[36] G. C. Calafiore and L. E. Ghaoui, "On distributionally robust chance-constrained linear programs," Journal of Optimization Theory and Applications, vol. 130, no. 1, pp. 1–22, 2006.

[37] R. Henrion and C. Strugarek, "Convexity of chance constraints with independent random variables," Computational Optimization and Applications, vol. 41, no. 2, pp. 263–276, 2008.

[38] P. Beraldi and A. Ruszczyński, "A branch and bound method for stochastic integer problems under probabilistic constraints," Optimization Methods and Software, vol. 17, no. 3, pp. 359–382, 2002.

[39] D. Dentcheva, A. Prékopa, and A. Ruszczynski, "Concavity and efficient points of discrete distributions in probabilistic programming," Mathematical programming, vol. 89, no. 1, pp. 55–77, 2000.

[40] A. Prékopa, "Dual method for the solution of a one-stage stochastic programming problem with random RHS obeying a discrete probability distribution," Zeitschrift für Operations Research, vol. 34, no. 6, pp. 441–461, 1990.

[41] L. Hewing and M. N. Zeilinger, "Stochastic model predictive control for linear systems using probabilistic reachable sets," in 2018 IEEE Conference on Decision and Control (CDC), 2018, pp. 5182–5188.

[42] S. Yan, P. Goulart, and M. Cannon, "Stochastic model predictive control with discounted probabilistic constraints," in 2018 European Control Conference (ECC), 2018, pp. 1003–1008.

[43] A. Nemirovski, "On safe tractable approximations of chance constraints," European Journal of Operational Research, vol. 219, no. 3, pp. 707–718, 2012.

[44] W. Chen, M. Sim, J. Sun, and C.-P. Teo, "From CVaR to uncertainty set: Implications in joint chance-constrained optimization," Operations Research, vol. 58, no. 2, pp. 470–485, 2010.

[45] I. Bremer, R. Henrion, and A. Möller, "Probabilistic constraints via SQP solver: application to a renewable energy management problem," Computational Management Science, vol. 12, no. 3, pp. 435–459, 2015.

[46] A. Jasour, N. S. Aybat, and C. M. Lagoa, "Semidefinite programming for chance constrained optimization over semialgebraic sets," SIAM Journal on Optimization, vol. 25, no. 3, pp. 1411–1440, 2015.

[47] J. B. Lasserre, "Representation of chance-constraints with strong asymptotic guarantees," IEEE Control Systems Letters, vol. 1, no. 1, pp. 50–55, 2017.

[48] S. Ahmed and W. Xie, "Relaxations and approximations of chance constraints under finite distributions," Mathematical Programming, vol. 170, no. 1, pp. 43–65, 2018.

[49] M. A. Lejeune, "Pattern-based modeling and solution of probabilistically constrained optimization problems," Operations research, vol. 60, no. 6, pp. 1356–1372, 2012.

[50] M. A. Lejeune and F. Margot, "Solving chance-constrained optimization problems with stochastic quadratic inequalities," Operations Research, vol. 64, no. 4, pp. 939–957, 2016.

[51] J. Luedtke and S. Ahmed, "A sample approximation approach for optimization with probabilistic constraints," SIAM Journal on Optimization, vol. 2, no. 19, pp. 674–699, 2008.

[52] 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.

[53] G. Calafiore and M. Campi, "The scenario approach to robust control design," IEEE Transactions on Automatic Control, vol. 51, no. 5, pp. 742–753, 2006.

[54] M. Campi and S. Garatti, "The exact feasibility of randomized solutions of robust convex programs," SIAM Journal of Optimization, vol. 19, pp. 1211—1230, 2008.

[55] G. Calafiore, "Random convex programs," SIAM Journal of Optimization, vol. 20, pp. 3427–3464, 2010.

[56] M. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality," Journal of Optimization Theory and Applications, vol. 148, pp. 257–280, 2011.

[57] V. Vapnik, Statistical Learning Theory. New York: John Wiley and Sons, 1998.

[58] M. Mammarella, V. Mirasierra, M. Lorenzen, T. Alamo, and F. Dabbene, "Chance constrained sets approximation: A probabilistic scaling approach – extended version," arXiv preprint arXiv:2101.06052, pp. 1–16, 2021.

[59] M. Herceg, M. Kvasnica, C. N. Jones, and M. Morari, "Multi-parametric toolbox 3.0," in 2013 European Control Conference (ECC). IEEE, 2013, pp. 502–510.

[60] F. Dabbene, C. Lagoa, and P. Shcherbakov, "On the complexity of randomized approximations of nonconvex sets," in 2010 IEEE International Symposium on Computer-Aided Control System Design. IEEE, 2010, pp. 1564–1569.

[61] V. T. H. Le, C. Stoica, T. Alamo, E. F. Camacho, and D. Dumur, Zonotopes: From Guaranteed State-estimation to Control. Wiley, 2013.

[62] F. Dabbene, D. Henrion, C. Lagoa, and P. Shcherbakov, "Randomized approximations of the image set of nonlinear mappings with applications to filtering," IFAC-PapersOnLine, vol. 48, no. 14, pp. 37–42, 2015.

[63] T. Alamo, J. M. Bravo, and E. F. Camacho, "Guaranteed state estimation by zonotopes," Automatica, vol. 41, no. 6, pp. 1035–1043, 2005.

[64] E. Gover and N. Krikorian, "Determinants and the volumes of parallelotopes and zonotopes," Linear Algebra and its Applications, vol. 433, no. 1, pp. 28–40, 2010.

[65] M. D. Buhmann, "Radial basis functions," Acta numerica, vol. 9, pp. 1–38, 2000.

[66] F. Dabbene, D. Henrion, and C. M. Lagoa, "Simple approximations of semialgebraic sets and their applications to control," Automatica, vol. 78, pp. 110 – 118, 2017.

[67] T. Alamo, R. Tempo, A. Luque, and D. Ramirez, "Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms," Automatica, vol. 52, pp. 160–172, 2015.