2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Learning for Control: a Bayesian Scenario Approach

Simone Garatti and Marco C. Campi

Abstract— The scenario approach is a data-driven method for uncertain optimization that in recent years has found many applications in systems and control. In this context, a decision is built from a sample of observations and the "risk" is the probability that the scenario decision meets a shortfall in a new, out-of-sample, case. This paper focuses on the "complexity of the solution" (as precisely defined in the paper) and studies the conditional probability distribution of the risk given the solution complexity. By a fundamental theoretical limitation (shown in the paper), no conditional results can be drawn without additional knowledge on uncertainty. This paper thus introduces a new perspective where prior knowledge can be incorporated and the main achievement is that strong conditional results can be established under very mild priors. This result allows for tight, a-posteriori, evaluations of the risk and improves the usability of the approach.

I. INTRODUCTION AND PROBLEM POSITION

The scenario approach is a sample-based method to deal with optimization problems under uncertainty, a condition which ever more often arises in systems and control design. In mathematical terms, we consider optimization problems with convex cost function $f(x)$ and uncertain convex constraint $x \in \mathcal{X}_\delta$, where $x \in \mathcal{X} \subseteq \mathbb{R}^d$ is the optimization variable and δ is a random element, defined over a probability space $(\Delta, \mathcal{D}, \mathbb{P})$, that is used to model uncertainty. All involved sets $\mathcal{X}$ and $\mathcal{X}_\delta$ are assumed to be convex. Given $N \geq d$ independent realizations $\delta_1, \delta_2, \ldots, \delta_N$ of δ, a scenario program is written as

$$SP_N : \quad \min_{x \in \mathcal{X}} f(x)$$
$$\text{subject to: } x \in \bigcap_{i=1,\ldots,N} \mathcal{X}_{\delta_i}, \quad (1)$$

that is, in the scenario program only finitely many constraints given by $\delta_1, \delta_2, \ldots, \delta_N$ are enforced and used to confine the choice of the optimization variable x. Scenario programs of the above type have been much studied over the past decade, [4], [9], [1], [10], [30], [14], [23], [34], [15], [11], [28]. They offer a general-purpose methodology that has proven useful in many contexts, including control systems design, [5], [12], [19], [29], [2], [21], system identification, [7], [18], machine learning, [6], [8], [24], [16], and quantitative finance, [25], [26], [27], [22] to mention but a few. Throughout this paper it is assumed that program (1) admits solution. If more than one solution exists, we assume that a solution is singled out by a convex rule, that is, the tie is broken by minimizing an

S. Garatti is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, piazza L. da Vinci 32, 20133 Milano, Italy. M.C. Campi is with the Department of Information Engineering, University of Brescia, via Branze 38, 25123 Brescia, Italy. Emails: `simone.garatti@polimi.it`, `marco.campi@unibs.it`.

additional convex function $t_1(x)$, and, possibly, other convex functions $t_2(x)$, $t_3(x)$, ... if the tie still occurs. An example of a tie-break function is the norm of x, $t_1(x) = \|x\|$. Another example is the lexicographic rule, which consists in minimizing the components of x in succession, i.e., $t_1(x) = x_1$, $t_2(x) = x_2$, ..., $t_d(x) = x_d$. After breaking the tie, the unique solution is denoted by x_N^*.

The $\delta_1, \delta_2, \ldots, \delta_N$ are the so-called scenarios and are interpreted as a record of empirical observations of the random variable δ coming from past experience. The idea underlying scenario optimization is that enforcing the satisfaction of the sole constraints $x \in \mathcal{X}_{\delta_i}$, $i = 1, \ldots, N$, guarantees robustness against the large part of the possible realizations of $\delta \in \Delta$. Moreover, considering only $\delta_1, \delta_2, \ldots, \delta_N$ lessens the difficulties that are inherent in managing the wealth of all the realizations of $\delta \in \Delta$, which is either not viable because of computational issues or even impossible because Δ and $\mathbb{P}$ are unknown to the user. The *generalization result* which ties $\delta_1, \delta_2, \ldots, \delta_N$ to $\delta \in \Delta$ is reviewed in the following as it provides a basis of analysis to better understand the results of the present contribution.

Start by introducing the following definition of violation.

Definition 1 (violation): Given an $x \in \mathcal{X}$, the *violation* of x is defined as $V(x) = \mathbb{P}\{\delta \in \Delta : x \notin \mathcal{X}_\delta\}$.　　★

$V(x)$, which is also called "risk", quantifies the probability with which a new randomly selected constraint is violated by x, and thereby quantifies the level of robustness of x against the uncertain constraint $x \in \mathcal{X}_\delta$. $V(x)$ also has an interpretation in terms of repeated experiments as follows. Consider the infinite product probability space $(\Delta^\infty, \mathcal{D}^\infty, \mathbb{P}^\infty)$. By the law of large numbers, one has

$$\lim_{M \to \infty} \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}_{x \notin \mathcal{X}_{\delta_{N+j}}} = V(x), \quad (2)$$

$\mathbb{P}^\infty$-almost surely. Hence, $V(x)$ is the long term average of times where x does not satisfy a sequence of independent constraints.

When x in $V(x)$ is replaced by the solution of the scenario program x_N^* (which is random given its dependence on $\delta_1, \ldots, \delta_N$), one obtains a random variable $V(x_N^*)$ and, in applications, one is interested to know how this random variable distributes to judge the robustness level of x_N^*, [9], [11]. One fundamental result [9] states that the distribution of $V(x_N^*)$ is always dominated by a Beta$(d, N - d + 1)$ distribution. That is, irrespective of $\mathbb{P}$ it holds that

$$\mathbb{P}^N\{V(x_N^*) \leq \epsilon\} \geq 1 - \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1 - \epsilon)^{N-i}. \quad (3)$$

This result is tight since (3) holds with equality for a whole class of problems, those called fully-supported in [9]. Moreover, being (3) valid for any $\mathbb{P}$ makes this result widely applicable to situations where knowledge on $\mathbb{P}$ beyond the sample $\delta_1, \delta_2, \ldots, \delta_N$ is scarce or unavailable.

An inspection of how the result (3) has been proven in [9] shows the central role played by the concept of support set.

Definition 2 (support set): A sub-sample $\delta_{i_1}, \ldots, \delta_{i_k}$ of the scenarios is said to be a support set if the solution obtained by enforcing the constraints associated to $\delta_{i_1}, \ldots, \delta_{i_k}$ only coincides with x_N^*, the solution with all the scenarios in place. ★

In [4], it is shown that the cardinality of the smallest support set (i.e. a support set with a minimal number of elements) never exceeds the number of optimization variables d, and this fact is used to set the upper limit to the distribution of $V(x_N^*)$ given in (3) which explicitly contains d. As d increases, the Beta distribution shifts to the right signifying that the desired event $\{V(x_N^*) \leq \epsilon\}$ becomes less probable. On the other hand, it is not rare that if one a-posteriori determines the smallest support set after that x_N^* has been obtained,[1] fewer scenarios are found in it than there are optimization variables. This is especially true for optimization problems in high dimensions where the gap between the cardinality of the smallest support set and d is often large (see e.g. [33], [32], [30], [14], [11], [13] for examples in various contexts). When this happens, it comes spontaneous to ask whether a better bound than (3) can be used, and particularly whether in (3) d can be substituted by the actual cardinality of the smallest support set, say h, and still obtain a valid bound conditionally on having a smallest support set with h elements. In formulas, this is written as

$$\mathbb{P}^N\{V(x_N^*) \leq \epsilon | s_N^* = h\} \geq 1 - \sum_{i=0}^{h-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}, \quad (4)$$

where s_N^* is a random variable that returns the cardinality of the smallest support set for (1) and conditioning is taken over the event where this cardinality equals h. It is a fact that this result is incorrect and the extremely negative result holds that the conditional probability on the left-hand side of (4) can be any small even for large ϵ. This is shown in the next example.

Example 1: Consider a problem where $x \in \mathbb{R}^2$, $c^T x = x_2$, and $\mathcal{X}_\delta$ is either V-shaped or U-shaped as depicted in Figure 1. More precisely, with probability $1-p$ $\mathcal{X}_\delta$ is V-shaped with vertex uniformly distributed along an horizontal segment; correspondingly, with probability p $\mathcal{X}_\delta$ is instead U-shaped with vertex uniformly distributed along a vertical segment. V-shaped constraints are all above U-shaped constraints. When N constraints are considered, it is clear that $s_N^* = 1$ happens only if either all the N constraints are U-shaped or all the seen V-shaped constraints have the same
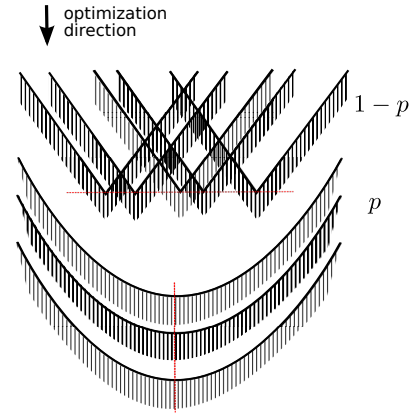

Fig. 1. A pictorial representation of V-shaped and U-shaped constraints. Shadings indicate the forbidden regions.

vertex and coincide. In both cases, the V-shaped constraints for all possible vertexes are, with the exception of one at most, all violated by x_N^*, and violation $V(x_N^*)$ is no smaller $1-p$. Thus,

$$\mathbb{P}^N\{V(x_N^*) \leq \epsilon | s_N^* = 1\} \leq \begin{cases} 0, & \epsilon < 1-p \\ 1, & \epsilon \geq 1-p \end{cases}. \quad (5)$$

By taking p close to 0, one sees that even for large ϵ the conditional probability on the left can be equal to zero. ★

The example proves that no conditional results can be affirmed without further knowledge on the optimization problem under consideration. Hence, we in this paper move on to a setting different from [9] where prior knowledge can be incorporated. The main contributions of this paper are:

- in Section II, we introduce a new Bayesian setup for scenario optimization that allows for the computation of the distribution of the violation conditionally on having a smallest support set of cardinality h;
- by elaborating upon the achievements of Section II, in Section III we show that strong conditional results can be established under very mild priors. This is of most importance for the usability of the theory. The computational advantages associated to the use of these priors are also discussed;
- in Section IV, we consider a flexible family of priors (the Dirichlet priors) that can accommodate a wide variety of situations. For this family of priors, explicit and handy expressions for the conditional distribution of the violation are provided;
- finally, in Section V a numerical example in control is presented that illustrates the effectiveness of the findings of this paper.

II. A BAYESIAN SETUP FOR SCENARIO OPTIMIZATION

To set the mathematical stage for the new Bayesian setup, consider a probability space $(\Theta, \mathcal{Q}, \pi)$ and let $\mathbb{P}_\vartheta$ be a transition probability function, [3], on $\Theta \times \mathcal{D}$ (recall that $\mathcal{D}$ is the σ-algebra of the probability space $(\Delta, \mathcal{D}, \mathbb{P})$ hosting the constraint parametrization δ):

---

[1] In the convex setup, active constraints form a support set and often this is also minimal. Hence, to compute the smallest support set, it is convenient to start by checking whether no elements can be further removed from the active constraints, which is computationally cheap. If this is not so, one can resort e.g. to the algorithm described in [13].

i. $\forall \vartheta \in \Theta$, $D \to \mathbb{P}_\vartheta(D)$ is a probability distribution;

ii. $\forall D \in \mathcal{D}$, $\vartheta \to \mathbb{P}_\vartheta(D)$ is $\mathcal{Q}$-measurable.

The interpretation is that, for any given $\vartheta \in \Theta$, $\mathbb{P}_\vartheta$ defines a specific uncertain optimization problem; the optimization problem at hand is only partially known though, and, to model this, ϑ is not taken deterministic, but it distributes according to π. The probability distribution π is regarded as the prior.

Next, consider the probability space $(\Delta^N \times \Theta, \mathcal{D}^N \otimes \mathcal{Q}, \mathrm{P})$ where P is the unique probability measure that extends the definition

$$\mathrm{P}(\tilde{D} \times Q) = \int_Q \mathbb{P}_\vartheta^N(\tilde{D}) \pi(\mathrm{d}\vartheta), \ \forall \tilde{D} \in \mathcal{D}^N, Q \in \mathcal{Q}, \quad (6)$$

see [31, Chapter 2, Section 9, Theorem 2]. $(\Delta^N \times \Theta, \mathcal{D}^N \otimes \mathcal{Q}, \mathrm{P})$ is the space that hosts a θ along with a sample of N independent constraints obtained from a problem where δ distributes according to $\mathbb{P}_\vartheta$. In this context, the definition of violation becomes $V_\vartheta(x) = \mathbb{P}_\vartheta\{\delta \in \Delta : x \notin \mathcal{X}_\delta\}$, and, for each ϑ, it admits the same long term interpretation as $V(x)$ in (2). $V_\vartheta(x_N^*)$ is the violation of the solution to (1) when $\delta_1, \dots, \delta_N$ is an i.i.d. sample from from $(\Delta, \mathcal{D}, \mathbb{P}_\vartheta)$ and ϑ is a realization from $(\Theta, \mathcal{Q}, \pi)$. As such, $V_\vartheta(x_N^*)$ is a random variable defined over the probability space $(\Delta^N \times \Theta, \mathcal{D}^N \otimes \mathcal{Q}, \mathrm{P})$. Also $s_N^* :=$ cardinality of the smallest support set for (1) is an integer random variable defined over the same space.

Our objective is to evaluate the distribution of $V_\vartheta(x_N^*)$ conditionally on observing a smallest support set with h elements, which is

$$F_V(\epsilon | s_N^* = h) = \mathrm{P}\{V_\vartheta(x_N^*) \le \epsilon | s_N^* = h\}.$$

$F_V(\epsilon | s_N^* = h)$ can in principle be calculated once π is given according the following formula:[2]

$$
\begin{aligned}
&F_V(\epsilon | s_N^* = h) \\
&= \ \mathrm{P}\{V_\vartheta(x_N^*) \le \epsilon | s_N^* = h\} \\
&= \ 1 - \mathrm{P}\{V_\vartheta(x_N^*) > \epsilon | s_N^* = h\} \\
&= \ 1 - \frac{\mathrm{P}\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h\}}{\mathrm{P}\{s_N^* = h\}} \\
&= \ [\text{use (6)}] \\
&= \ 1 - \frac{\int_\Theta \mathbb{P}_\vartheta^N\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h\}\pi(\mathrm{d}\vartheta)}{\int_\Theta \mathbb{P}_\vartheta^N\{s_N^* = h\}\pi(\mathrm{d}\vartheta)}. \quad (7)
\end{aligned}
$$

Unfortunately, formula (7) has to be regarded as a theoretical contribution only, with no practical value for the actual calculation of $F_V(\epsilon | s_N^* = h)$. The reason is twofold:

1. to model real applications where one has a substantial lack of knowledge on the problem at hand, Θ need to be a high dimensional, dramatically complicated, space (in most situations, Θ is even an infinite dimensional space of probability distributions!). In this case, the computation of the integrals appearing in (7) is demanding, if not impossible;

2. more intrinsically, when Θ is so complicated, a sensible choice for the prior π can be hard, if not impossible, to make (e.g. when Θ is an infinite dimensional space of probability distributions, it is even not always clear how to endow this space with a π). Hence, in these situations one lacks the essential information for calculating (7).

In the next section we shall show that tight and useful bounds for $F_V(\epsilon | s_N^* = h)$ can be drawn based on partial knowledge on π. Precisely, let π' be the distribution of $\mathbb{P}_\vartheta^N\{s_N^* = 0\}, \mathbb{P}_\vartheta^N\{s_N^* = 1\}, \dots, \mathbb{P}_\vartheta^N\{s_N^* = d\}$ induced by π. This is a finite-dimensional distribution over the simplex in dimension $d + 1$ that describes prior knowledge on the distribution of the cardinality of the smallest support set. Our main result is that $F_V(\epsilon | s_N^* = h)$ can be effectively bounded based on π' only. This result is formalized and precisely stated in the next section.

III. Evaluation of the a-posteriori conditional distribution of the violation

A proper evaluation of $F_V(\epsilon | s_N^* = h)$ rests upon a suitable upper-bound of the integrand $\mathbb{P}_\vartheta^N\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h\}$ at the numerator of (7). We have the following lemma.

Lemma 1: For all $\vartheta \in \Theta$, it holds that

$$\mathbb{P}_\vartheta^N\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h\} \le b_h(\epsilon), \quad (8)$$

where

$$b_h(\epsilon) = \min\left\{\binom{N}{h}(1-\epsilon)^{N-h}, 1\right\}. \quad (9)$$

Proof: Suppose first that not only the cardinality of the smallest support set is h, but also that it is formed by the first h scenarios $\delta_1, \dots, \delta_h$. All the $N - h$ constraints associated with scenarios that are not in the smallest support set must be satisfied by the solution x_N^*, which is determined by $\delta_1, \dots, \delta_h$ only. If $V_\vartheta(x_N^*) > \epsilon$, this means that $\delta_{h+1}, \dots, \delta_N$ must belong to an event whose probability is no more than $1 - \epsilon$, and this happens with a probability that is no more than $(1-\epsilon)^{N-h}$ since the constraints are independent of each other. Hence, $\mathbb{P}_\vartheta^N\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h \wedge$ the smallest support set is formed by the first h constraints$\} \le (1-\epsilon)^{N-h}$. Now, let us vary which constraints belong to the smallest support set. Summing over all the $\binom{N}{h}$ possible choices of h scenarios out of N, we obtain:

$$
\begin{aligned}
&\mathbb{P}_\vartheta^N\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h\} \\
&\le \ \sum_{i=1}^{\binom{N}{h}} \mathbb{P}_\vartheta^N\Big\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h \wedge \text{ smallest support} \\
&\qquad\qquad \text{set is equal to the } i\text{-th group of } h \text{ constraints}\Big\} \\
&\le \ \binom{N}{h}(1-\epsilon)^{N-h}.
\end{aligned}
$$

This last inequality, along with the fact that a probability cannot be greater than one, establishes (8) and (9). ∎

To properly bound the integrand at the numerator of equation (7), we shall also use the trivial inequality

$$\mathbb{P}_\vartheta^N\{V_\vartheta(x_N^*) > \epsilon \wedge s_N^* = h\} \le \mathbb{P}_\vartheta^N\{s_N^* = h\},$$

[2]In the derivation it is assumed that $\mathrm{P}\{s_N^* = h\} > 0$; if not, $F_V(\epsilon | s_N^* = h)$ can be arbitrarily defined.

1774

which together with (8) gives

$$\mathbb{P}^N_\vartheta\{V_\vartheta(x^*_N) > \epsilon \wedge s^*_N = h\} \tag{10}$$
$$\leq \begin{cases} \mathbb{P}^N_\vartheta\{s^*_N = h\}, & \text{if } \mathbb{P}^N_\vartheta\{s^*_N = h\} \leq b_h(\epsilon) \\ b_h(\epsilon), & \text{if } \mathbb{P}^N_\vartheta\{s^*_N = h\} > b_h(\epsilon) \end{cases}.$$

Using (10) in (7) gives

$$F_V(\epsilon|s^*_N = h)$$
$$\geq 1 - \frac{\int_{\{\vartheta:\, \mathbb{P}^N_\vartheta\{s^*_N=h\}\leq b_h(\epsilon)\}} \mathbb{P}^N_\vartheta\{s^*_N = h\}\pi(\mathrm{d}\vartheta)}{\int_\Theta \mathbb{P}^N_\vartheta\{s^*_N = h\}\pi(\mathrm{d}\vartheta)}$$
$$- \frac{\int_{\{\vartheta:\, \mathbb{P}^N_\vartheta\{s^*_N=h\}> b_h(\epsilon)\}} b_h(\epsilon)\pi(\mathrm{d}\vartheta)}{\int_\Theta \mathbb{P}^N_\vartheta\{s^*_N = h\}\pi(\mathrm{d}\vartheta)}$$
$$= 1 - \frac{\int_{\{\vartheta:\, \mathbb{P}^N_\vartheta\{s^*_N=h\}\leq b_h(\epsilon)\}} \mathbb{P}^N_\vartheta\{s^*_N = h\}\pi(\mathrm{d}\vartheta)}{\int_\Theta \mathbb{P}^N_\vartheta\{s^*_N = h\}\pi(\mathrm{d}\vartheta)}$$
$$- \frac{b_h(\epsilon)\pi\{\mathbb{P}^N_\vartheta\{s^*_N = h\} > b_h(\epsilon)\}}{\int_\Theta \mathbb{P}^N_\vartheta\{s^*_N = h\}\pi(\mathrm{d}\vartheta)}. \tag{11}$$

Note that in (11) all the integrands depend on ϑ via $\mathbb{P}^N_\vartheta\{s^*_N = h\}$ only. Define, thus, $p_k := \mathbb{P}^N_\vartheta\{s^*_N = k\}$, $k = 0, 1, \ldots, d$, and let $\boldsymbol{p} = (p_0, p_1, \ldots, p_d) \in S$ where S is the simplex in $\mathbb{R}^{d+1}$ (i.e., $\sum_{k=0}^d p_k = 1$, $p_k \geq 0$, $k = 0, 1, \ldots, d$). $\boldsymbol{p}$ is a random variable because of its dependence on ϑ. Letting π' be the probability measure of $\boldsymbol{p}$ induced by π, a change of variables in (11) yields the sought bound on $F_V(\epsilon|s^*_N = h)$ that depends on π' only.

Theorem 1: It holds that

$$F_V(\epsilon|s^*_N = h) \geq \tag{12}$$
$$1 - \frac{\int_{\{p_h \leq b_h(\epsilon)\}} p_h \pi'(\mathrm{d}\boldsymbol{p}) + b_h(\epsilon)\pi'\{p_h > b_h(\epsilon)\}}{\int_S p_h \pi'(\mathrm{d}\boldsymbol{p})}.$$

The bound (12) has a clear computational advantage over (7) since the integrals appearing in (12) are all defined over a finite $(d+1)$-dimensional space and can be computed via standard techniques in most cases. Besides the computational aspects, the main thrust of Theorem 1 is that the inference process to obtain an evaluation of $F_V(\epsilon|s^*_N = h)$ can be pursued by directly choosing the prior π' rather than π. π' is a probability distribution over a finite-dimensional space and represents prior knowledge on the distribution of the cardinality of the smallest support set. π' is much simpler to obtain than π and it can also be learned by experience.

IV. DIRICHLET PRIORS

Given that $\boldsymbol{p}$ is defined over a simplex, a rather natural family of prior distributions for $\boldsymbol{p}$ are the so-called Dirichlet priors. Specifically,

$$\pi' = \mathrm{Dir}(\alpha_0, \alpha_1, \ldots, \alpha_d),$$

where $\mathrm{Dir}(\alpha_0, \alpha_1, \ldots, \alpha_d)$ is the probability distribution obtained by endowing $p_1, \ldots, p_d$ with the density

$$\frac{\Gamma\left(\sum_{k=0}^d \alpha_k\right)}{\prod_{k=0}^d \Gamma(\alpha_k)} \left(1 - \sum_{k=1}^d p_k\right)^{\alpha_0 - 1} \prod_{k=1}^d p_k^{\alpha_k - 1}$$

over the support $\sum_{i=1}^d p_k \leq 1$, $p_k \geq 0$, $k = 1, \ldots, d$, and letting $p_0 = 1 - \sum_{k=1}^d p_k$. In this formula, $\Gamma(\cdot)$ is the Gamma function and $\alpha_0 > 0$, $\alpha_1 > 0$, $\ldots$, $\alpha_d > 0$ are degrees of freedom. By tuning them, an extremely rich variety of priors
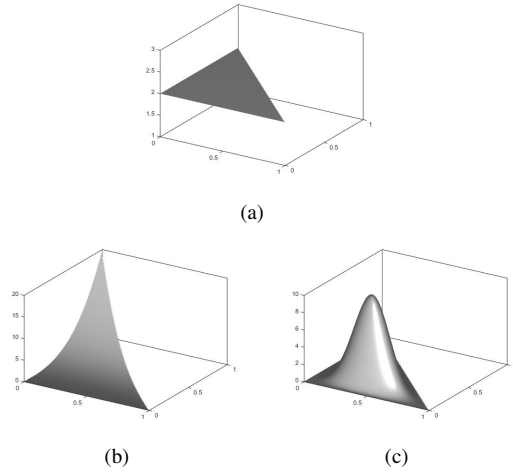

(a)

(b)

(c)

Fig. 2. Density of p_1, p_2 for a Dirichlet distribution with: (a) $\alpha_0 = \alpha_1 = \alpha_2 = 1$; (b) $\alpha_0 = \alpha_1 = 1, \alpha_2 = 4$; (c) $\alpha_0 = \alpha_1 = \alpha_2 = 10$.

is obtained, which can accommodate many cases of interest. Figure 2 displays some of the achievable priors for some choices of the degrees of freedom when $d = 2$.

In this section, we give an explicit expression for the bound on $F_V(\epsilon|s^*_N = h)$ given in (12) when Dirichlet priors are used. To this purpose, it is useful to first recall a well known result on Dirichlet distributions, namely, that the marginal probability distribution of p_h, say π'_h, is a Beta distribution. To be precise, $\pi'_h = \mathrm{Beta}(\alpha_h, \sum_{k \neq h} \alpha_k)$, where $\mathrm{Beta}(\alpha_h, \sum_{k \neq h} \alpha_k)$ is the distribution corresponding to the density

$$\frac{\Gamma\left(\sum_{k=0}^d \alpha_k\right)}{\Gamma(\alpha_h)\Gamma(\sum_{k \neq h} \alpha_k)} p_h^{\alpha_h - 1}(1 - p_h)^{\sum_{k \neq h} \alpha_k - 1}$$

over the support $p_h \in [0, 1]$. $\mathrm{Beta}(a, b)$ is another well known distribution in probability and statistics. Its main properties are that its mean is equal to $a/(a + b)$, and that its cumulative distribution function

$$\mathrm{I}_x(a, b) = \int_0^x \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1 - t)^{b-1}\mathrm{d}t$$

is the so-called regularized incomplete Beta function. The value of $\mathrm{I}_x(a, b)$ for any x, a and b can be efficiently computed in most of the available scientific computing environments; for example, through the command `betainc(x,a,b)` in MATLAB.

We are now ready to give the main result of this section, providing an handy and immediate to implement expression for the bound to $F_V(\epsilon|s^*_N = h)$ in (12) when a Dirichlet prior is used.

Theorem 2: If $\pi' = \mathrm{Dir}(\alpha_0, \alpha_1, \ldots, \alpha_d)$, then it holds

1775

that

$$F_V(\epsilon|s_N^* = h) \geq 1 - \left[I_{b_h(\epsilon)}\left(\alpha_h + 1, \sum_{k \neq h} \alpha_k\right) \right. \tag{13}$$

$$\left. + b_h(\epsilon) \cdot \frac{\sum_{k=0}^d \alpha_k}{\alpha_h}\left(1 - I_{b_h(\epsilon)}\left(\alpha_h, \sum_{k \neq h} \alpha_k\right)\right) \right].$$

Proof: The result is simply obtained by evaluating the integrals appearing in the right-hand side of (12). The integral at the denominator is easily recognized to be equal to the mean of $\text{Beta}(\alpha_h, \sum_{k \neq h} \alpha_k)$:

$$\int_S p_h \pi'(\mathrm{d}\boldsymbol{p}) = \int_0^1 p_h \pi'_h(\mathrm{d}p_h) = \frac{\alpha_h}{\sum_{k=0}^d \alpha_k}. \tag{14}$$

As for the terms at the numerator, instead, we have that

$$\begin{aligned} \pi'\{p_h > b_h(\epsilon)\} &= \pi'_h\{p_h > b_h(\epsilon)\} \\ &= \int_{b_h(\epsilon)}^1 \frac{\Gamma\left(\sum_{k=0}^d \alpha_k\right)}{\Gamma(\alpha_h)\Gamma(\sum_{k \neq h} \alpha_k)} p_h^{\alpha_h - 1} \times \\ &\quad \times (1 - p_h)^{\sum_{k \neq h} \alpha_k - 1} \mathrm{d}p_h \\ &= 1 - I_{b_h(\epsilon)}\left(\alpha_h, \sum_{k \neq h} \alpha_k\right), \end{aligned} \tag{15}$$

and that

$$\begin{aligned} &\int_{\{p_h \leq b_h(\epsilon)\}} p_h \pi'(\mathrm{d}\boldsymbol{p}) \\ &= \int_{\{p_h \leq b_h(\epsilon)\}} p_h \pi'_h(\mathrm{d}p_h) \\ &= \int_0^{b_h(\epsilon)} \frac{\Gamma\left(\sum_{k=0}^d \alpha_k\right)}{\Gamma(\alpha_h)\Gamma(\sum_{k \neq h} \alpha_k)} p_h^{\alpha_h} \times \\ &\quad \times (1 - p_h)^{\sum_{k \neq h} \alpha_k - 1} \mathrm{d}p_h \\ &= \text{since } [\Gamma(x+1) = x\Gamma(x)] \\ &= \frac{\alpha_h}{\sum_{k=0}^d \alpha_k} \cdot I_{b_h(\epsilon)}\left(\alpha_h + 1, \sum_{k \neq h} \alpha_k\right). \end{aligned} \tag{16}$$

Using (14), (15), and (16) in (12), the inequality (13) is eventually obtained. ∎

V. NUMERICAL EXAMPLE

Inspired by [17], we consider a finite-horizon optimal control problem arising in MPC design for the mechanical system depicted in Figure 3. The system is composed by
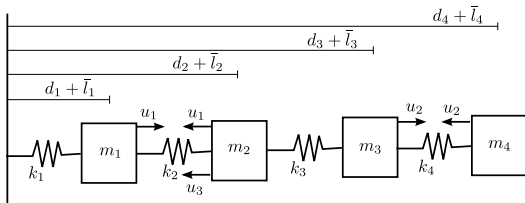

Fig. 3. The studied mechanical system.

four masses connected by springs and its state is a 8-dimensional vector $\xi = [d_1\ d_2\ d_3\ d_4\ \dot{d}_1\ \dot{d}_2\ \dot{d}_3\ \dot{d}_4]^T$, where d_1, d_2, d_3, d_4 are the mass displacements from the nominal positions $\bar{l}_1, \bar{l}_2, \bar{l}_3, \bar{l}_4$, while $\dot{d}_1, \dot{d}_2, \dot{d}_3, \dot{d}_4$ are the displacements' derivatives. The control input is instead $u = [u_1, u_2, u_3]^T$, where u_1, u_2, u_3 are forces acting on the masses as shown in Figure 3. All masses and stiffness constants are equal to 1.

In the example, the control action is kept constant over the sampling period and we work with the resulting discretized model of the system. This writes as

$$\xi_{t+1} = A\xi_t + Bu_t + Dw_t,$$

where A and B are suitable matrices and the term Dw_t is introduced to model a stochastic disturbance. We take

$$D = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T,$$

which means that the disturbance affects the fourth mass only, while w_t is a bi-variate noise.[3] The system is supposed to be at rest at $t = 0$.

The disturbance can be reconstructed from the state according to $w_t = D^\dagger(\xi_{t+1} - A\xi_t - Bu_t)$, where $D^\dagger$ is the pseudo-inverse of D. This motivates the following parametrization of the control action as an affine function of the disturbance:

$$u_t = \gamma_t + \sum_{\tau=0}^{t-1} \theta_{t,\tau} w_\tau, \tag{17}$$

where $\gamma_t \in \mathbb{R}^3$ and $\theta_{t,\tau} \in \mathbb{R}^{3 \times 2}$ are design parameters. Parametrization (17) was first proposed in [20] and has the great advantage of making u and ξ linear in the design parameters. After collecting $N = 500$ realizations $\delta_i = [w_0^{(i)}\ w_1^{(i)} \ldots w_4^{(i)}]$, $i = 1, \ldots, N$, of the disturbance along a horizon of 5 time-instants, γ_t and $\theta_{t,\tau}$ are designed according to the following scenario program:

$$\min_{\gamma_t, \theta_{t,\tau}} \quad \sum_{t=0}^4 \left[\|\gamma_t\|_2^2 + \sum_{\tau=0}^{t-1} \|\theta_{t,\tau}\|_F^2 \right] \tag{18}$$

$$\text{s.t.} \quad \sup_{t=1,\ldots,5} \left\| \begin{array}{c} d_{1,t}^{(i)} \\ d_{2,t}^{(i)} - d_{1,t}^{(i)} \\ d_{3,t}^{(i)} - d_{2,t}^{(i)} \\ d_{4,t}^{(i)} - d_{3,t}^{(i)} \end{array} \right\|_\infty \leq 1, \ i = 1, \ldots, N.$$

In (18), $\|\cdot\|_F$ denotes the Frobenius norm and $d_{1,t}^{(i)}$, $d_{2,t}^{(i)}$, $d_{3,t}^{(i)}$, $d_{4,t}^{(i)}$ are the first four components of the state corresponding to the i-th realization of the disturbance. The interpretation is that the cost accounts for the magnitude of control actions, while the constraints impose that the spring deformations are kept within a safe range.

The violation of the solution to (18) can be here interpreted as a measure of robustness as for the satisfaction of the deformation limits when new realizations of the disturbance are

---

[3]Though immaterial for the further development of the example, we report here for the sake of completeness that w_t is a white noise, with independent components $w_{1,t}, w_{2,t}$, each uniformly distributed over $[-0.6, 0.6]$.

faced. Evaluating the violation is of paramount importance for a reliable usage of the designed control policy. In (18), the total number of optimization variables is equal to $d = 75$, but after solving it we observe that the cardinality of the smallest support set s_{500}^* is just 11. We thus resort to Theorem 2 to assess $F_V(\epsilon|s_{500}^* = 11)$.

Take $\pi' = \mathrm{Dir}(\alpha_0, \alpha_1, \ldots, \alpha_{75})$ with $\alpha_0 = \cdots = \alpha_{75} = 1$, which corresponds to a flat prior giving the same likeliness to all possible situations. The resulting marginal π'_{11} is depicted in Figure 4.a (solid line), while the bound to $F_V(\epsilon|s_{500}^* = 11)$ as obtained from Theorem 2 is in Figure 4.b (solid line again). As it appears the conditional distribution
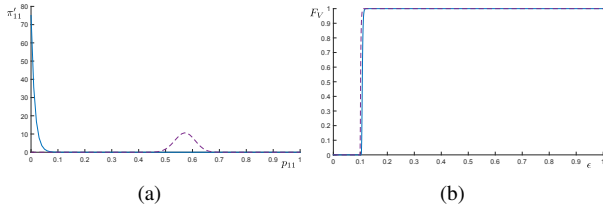

Fig. 4. (a) π'_{11}; (b) bound for $F_V(\epsilon|s_{500}^* = 11)$. Solid blue line: $\alpha_k = 1$, $k = 0, \ldots, 75$; dashed purple line: $\alpha_k = 1$, $k \neq 11$, $\alpha_{11} = 100$.

is concentrated and allows one to draw the conclusion that, for the specific case at hand with $s_{500}^* = 11$, it is very likely that the violation is below 10%.

Suppose now that $\alpha_k = 1$ for $k \neq 11$ and $\alpha_{11} = 100$, which amounts to having a strong beliefs that $s_{500}^* = 11$ has a high chance to happen. Interestingly, though the marginal is very different from the previous case (see Figure 4.a, dashed line), the bound to $F_V(\epsilon|s_{500}^* = 11)$ returned by Theorem 2 is very similar to the previous one (Figure 4.b, dashed line). This heuristically shows that the conditional distribution of the violation is little sensitive to the prior π'.

REFERENCES

[1] T. Alamo, R. Tempo, and E.F. Camacho. A randomized strategy for probabilistic solutions of uncertain feasibility and optimization problems. *IEEE Transactions on Automatic Control*, 54(11):2545–2559, 2009.

[2] T. Alamo, R. Tempo, A. Luque, and D. R. Ramirez. Randomized methods for design of uncertain systems: sample complexity and sequential algorithms. *Automatica*, 51:160–172, 2015.

[3] J.R. Barra. *Mathematical Basis of Statistics*. Academic Press, New York, NY, USA, 1981.

[4] G.C. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.

[5] G.C. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[6] M.C. Campi. Classification with guaranteed probability of error. *Machine Learning*, 80:63–84, 2010.

[7] M.C. Campi, G. Calafiore, and S. Garatti. Interval predictor models: identification and reliability. *Automatica*, 45(2):382–392, 2009.

[8] M.C. Campi and A. Carè. Random convex programs with l_1-regularization: sparsity and generalization. *SIAM Journal on Control and Optimization*, 51(5):3532–3557, 2013.

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

[10] M.C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011.

[11] M.C. Campi and S. Garatti. Wait-and-judge scenario optimization. *Mathematical Programming*, 167(1):155–189, 2018.

[12] M.C. Campi, S. Garatti, and M. Prandini. The scenario approach for systems and control design. *Annual Reviews in Control*, 33(2):149 – 157, 2009.

[13] M.C. Campi, S. Garatti, and F.A. Ramponi. A general scenario theory for non-convex optimization and decision making. *IEEE Transactions on Automatic Control*, 63(12):4067–4078, 2018.

[14] A. Carè, S. Garatti, and M.C. Campi. FAST - Fast Algorithm for the Scenario Technique. *Operations Research*, 62(3):662–671, 2014.

[15] A. Carè, S. Garatti, and M.C. Campi. Scenario min-max optimization and the risk of empirical costs. *SIAM Journal on Optimization*, 25(4):2061–2080, 2015.

[16] A. Caré, F.A. Ramponi, and M.C. Campi. A new classification algorithm with guaranteed sensitivity and specificity for medical applications. *IEEE Control Systems Letters*, 2(3):393–398, 2018.

[17] E. Cinquemani, M. Agarwal, D. Chatterjee, and J. Lygeros. Convexity and convex approximations of discrete-time stochastic control problems with constraints. *Automatica*, 47(9):2082–2087, 2011.

[18] L.G. Crespo, S.P. Kenny, and D.P. Giesy. Random predictor models for rigorous uncertainty quantification. *International Journal for Uncertainty Quantification*, 5(5):469–489, 2015.

[19] S. Garatti and M.C. Campi. Modulating robustness in control design: principles and algorithms. *IEEE Control Systems Magazine*, 33(2):36–51, 2013.

[20] P.J. Goulart, E.C. Kerrigan, and J.M. Maciejowski. Optimization over state feedback policies for robust control with constraints. *Automatica*, 42(4):523–533, 2006.

[21] S. Grammatico, X. Zhang, K. Margellos, P.J. Goulart, and J. Lygeros. A scenario approach for non-convex control design. *IEEE Transactions on Automatic Control*, 61(2):334–345, 2016.

[22] L.J. Hong, Z. Hu, and G. Liu. Monte Carlo methods for value-at-risk and conditional value-at-risk: a review. *ACM Transactions on Modeling and Computer Simulation*, 24(4):22:1–22:37, 2014.

[23] K. Margellos, P.J. Goulart, and J. Lygeros. On the road between robust optimization and the scenario approach for chance constrained optimization problems. *IEEE Transactions on Automatic Control*, 59(8):2258–2263, 2014.

[24] 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*, 60(10):2716–2721, 2015.

[25] 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*, 142(2):399–416, 2009.

[26] B.K. Pagnoncelli, D. Reich, and M.C. Campi. Risk-return trade-off with the scenario approach in practice: A case study in portfolio selection. *Journal of Optimization Theory and Applications*, 155(2):707–722, 2012.

[27] B.K. Pagnoncelli and S. Vanduffel. A provisioning problem with stochastic payments. *European Journal of Operational Research*, 221(2):445–453, 2012.

[28] F.A. Ramponi. Consistency of the scenario approach. *SIAM Journal on Optimization*, 28:135–162, 2018.

[29] G. Schildbach, L. Fagiano, C. Frei, and M. Morari. The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations. *Automatica*, 50(12):3009–3018, 2014.

[30] G. Schildbach, L. Fagiano, and M. Morari. Randomized solutions to convex programs with multiple chance constraints. *SIAM Journal on Optimization*, 23(4):2479–2501, 2013.

[31] A.N. Shiryaev. *Probability-1*. Springer, New York, NY, USA, 2016.

[32] J.S. Welsh and H. Kong. Robust experiment design through randomisation with chance constraints. In *Proceedings of the 18th IFAC World Congress*, Milan, Italy, 2011.

[33] J.S. Welsh and C.R. Rojas. A scenario based approach to robust experiment design. In *Proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.

[34] X. Zhang, S. Grammatico, G. Schildbach, P.J. Goulart, and J. Lygeros. On the sample size of random convex programs with structured dependence on the uncertainty. *Automatica*, 60:182–188, 2015.