2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Scenario optimization with constraint relaxation in a non-convex setup: a flexible and general framework for data-driven design

Simone Garatti and Marco C. Campi

***Abstract*— The scenario approach, originally developed as a computational tool for robust problems, has through the years developed into a solid, general, framework for data-driven decision making and design. One main driving force that has fostered this process has certainly been the increasing generality of the considered schemes. In this paper, we move a further step forward in this process. By leveraging some recent results in the wake of the so-called wait-and-judge paradigm, we fully develop a scheme for scenario optimization with constraint relaxation in a non-convex setup, so greatly expanding previous achievements valid under a convexity assumption. We show that a purely data-driven, and yet tight and informative, quantification of the solution robustness is possible regardless of the mechanism through which uncertainty is generated. The generality of this new non-convex setup provides an extremely versatile scheme for data-driven design that can be applied to a variety of problems ranging from mixed-integer optimization to design in abstract spaces.**

I. Introduction

The scenario approach, [1], [2], is a framework to do data-driven design that has received increasing recognition by the systems and control community in recent years, see [3], [4], [5], [6], [7], [8], [9], [10], [11] among many theoretical contributions. Denoting by δ the vector that contains all the uncertain elements in the problem, the scenario approach moves from the assumption that one has at his disposal N observations $\delta_1, \dots, \delta_N$ of the variable δ, the so-called scenarios, modeled as i.i.d. draws from a probability space $(\Delta, \mathcal{D}, \mathbb{P})$. Probability $\mathbb{P}$ is meant to describe the mechanism that generates δ, but, as is typical in complex problems, $\mathbb{P}$ is not known or only imprecisely known at the user end, so that $\mathbb{P}$ cannot be used for design purposes. In this context, the goal of the scenario approach is that of providing design schemes that map $\delta_1, \dots, \delta_N$ into a decision that empirically achieves certain objectives while also offering theoretical tools by which the user can rigorously evaluate the robustness of the design against other occurrences of the uncertainty, beyond the observed scenarios.

A quite flexible scenario scheme that can accommodate numerous problems of interest is scenario optimization with constraints relaxation. Letting $x \in \mathcal{X}$ be the design variable, scenario optimization with constraints relaxation amounts to solve an optimization program of the form

$$\begin{aligned} \min_{x \in \mathcal{X}, \xi_i \geq 0} \quad & c(x) + \rho \sum_{i=1}^{N} \xi_i \\ \text{subject to:} \quad & f(x, \delta_i) \leq \xi_i, \quad i = 1, \dots, N. \end{aligned} \tag{1}$$

The solution to (1) is denoted by $x_N^*, \xi_{N,i}^*$, $i = 1, \dots, N$. The interpretation is as follows. The problem is formulated according to two ingredients: a cost function $c(x)$, measuring the intrinsic quality of a decision x, and $f(x, \delta)$, which is an indicator of the interaction of x with the environment and, hence, with the uncertain element δ. The minimization of $c(x)$ has to be regarded as a primary goal but, in order to achieve a meaningful design, one has also to account for the value of $f(x, \delta)$ over the scenarios as representatives of the possible occurrences of uncertainty. To set the problem, one identifies a threshold of satisfaction for the function $f(x, \delta)$ – which, by a suitable shifting of $f(x, \delta)$, is made to coincide with the value 0 without any loss of generality – leading to constraints of the type $f(x, \delta_i) \leq 0$. However, to avoid excessive stiffness and introduce more flexibility, scenario optimization with constraints relaxation admits that some constraints can be violated. This is achieved by the introduction of the variables ξ_i, $i = 1, \dots, N$, and relaxing the constraints to $f(x, \delta_i) \leq \xi_i$ generates a penalty $\rho \sum_{i=1}^{N} \xi_i$, which is added to $c(x)$, where ρ must be seen as a tuning knob at the user's disposal to modulate cost against constraint violation. When $\rho \to \infty$, one tends to robust optimization where no violation of the scenario constraints ia allowed, while $\rho = 0$ corresponds to the unconstrained minimization of $c(x)$. By selecting various values of ρ between these two extremes, the user can also explore competing solutions to chose from. This setup is extremely general and finds application to machine learning problems as well as to management and finance. An example in control is provided at the end of this paper.

The two fundamental indicators of the quality of x_N^* are given by the cost $c(x_N^*)$ and by the risk $V(x_N^*)$, which is obtained by plugging x_N^* into function $V(x)$ defined as

$$V(x) := \mathbb{P}\{\delta : \ f(x, \delta) > 0\}.$$

It is important to observe that $c(x_N^*)$ is immediately known as an outcome of the optimization procedure, but $V(x_N^*)$ cannot be evaluated from its definition because it depends on the unknown probability $\mathbb{P}$. While the user has an empirical indication of $V(x_N^*)$ through $(1/N) \sum_{i=1}^{N} \mathbf{1}_{\xi_{i,N}^* > 0}$ ($\mathbf{1}_A$ is the indicator function of set A), this empirical evaluation, as one can imagine, is a biased estimator of $V(x_N^*)$, and

S. Garatti is with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, piazza L. da Vinci 32, 20133 Milano, Italy. Email: simone.garatti@polimi.it.

M.C. Campi is with the Department of Information Engineering, University of Brescia, via Branze 38, 25123 Brescia, Italy. Email: marco.campi@unibs.it.

The research of Simone Garatti is is supported by PNRR-PE-AI FAIR project funded by the NextGeneration EU program.

indeed it can be severely misleading because the solution has a tendency to steer towards regions of low empirical violation that is not paired by an equally low real violation. The key problem addressed by the scenario theory is to obtain a trustworthy quantification of $V(x_N^*)$ from the available information, i.e., the scenarios $\delta_1, \dots, \delta_N$. For the decision scheme in (1), this problem was first addressed in [12], [13]. These papers, however, assume that (i) $c(x)$ and $f(x, \delta)$ are convex functions of x; and (ii) constraints do not accumulate, i.e. for any given $\bar{x}$ it must be that $\mathbb{P}\{f(\bar{x}, \delta) = 0\} = 0$. Condition (i) severely limits the class of problems under consideration, for instance it excludes mixed-integer problems, which are becoming ever more important in control to deal with cyber-physical systems. Condition (ii), instead, is mild by it ultimately depends on $\mathbb{P}$ and thus requires some prior knowledge on this probability, which in many cases is not available.

The main contribution of the present paper is to get rid of the assumptions on $c(x)$, $f(x, \delta)$ and $\mathbb{P}$, and show the rigorous validity of an informative bound on $V(x_N^*)$ that holds true in this general context. Towards this goal, we leverage an achievement of [14] that has identified a large class of decision schemes to which certain results of the scenario theory can be applied. In fact, our result in this paper will be proved by showing that optimization with constraint relaxation does belong to this class. As a second contribution, we also provide a bound on the probability that $f(x_N^*, \delta)$ exceeds a generic level $\ell \leq 0$ (as opposed to just $\ell = 0$), which provides additional information about the quality of the solution.

II. MAIN RESULTS

We start by clarifying a little more the mathematical setup. In the present paper, $\mathcal{X}$ is a completely generic space (no structure, like, e.g., that of vector space, is required) and $c(x)$ is any function from $\mathcal{X}$ to $\mathbb{R}$. Likewise, $(\Delta, \mathcal{D}, \mathbb{P})$ is a generic probability space and $f(x, \delta)$ is any function from $\mathcal{X} \times \Delta$ to $\mathbb{R}$. The only assumption we make is about the existence of the solution to (1). In this respect, note that because of the presence of the ξ_i's, problem (1) is never infeasible (for any $x \in \mathcal{X}$, one can just take large enough values of the variables ξ_i to satisfy all inequalities $f(x, \delta_i) \leq \xi_i$); nonetheless, the solution can still not exist because it may "drift" indefinitely in one direction (this happens when for any given point in $\mathcal{X} \times \mathbb{R}^N$ one can find another point in $\mathcal{X} \times \mathbb{R}^N$ that improves the performance). We assume that, for every $N \geq 0$ and for every choice of $\delta_1, \dots, \delta_N$, the $\min$ in (1) is attained in at least one point of the feasibility domain (for $N = 0$, x_0^* is meant to be the unconstrained minimizer of $c(x)$). In case of multiple minimizers, the solution $x_N^*, \{\xi_{N,i}^*\}_{i=1}^N$ is singled out by a rule of preference in the domain $\mathcal{X}$.[1]

A. Evaluation of the risk

We are now ready to present our first result. In the following, $\mathbb{P}^N$ is the probability distribution for $(\delta_1, \dots, \delta_N)$ and it is a product probability because the scenarios are independent draws.

Theorem 1: Given a confidence parameter $\beta \in (0, 1)$, for any $k = 0, 1, \dots, N-1$ consider the polynomial equation in the v variable

$$\binom{N}{k}(1-v)^{N-k} - \frac{\beta}{N}\sum_{m=k}^{N-1}\binom{m}{k}(1-v)^{m-k} = 0,$$

and let $\epsilon(k)$ be the unique solution of this equation over the interval $(0, 1)$.[2] Also define $\epsilon(N) = 1$. It holds that

$$\mathbb{P}^N\{V(x_N^*) > \epsilon(s_N^*)\} \leq \beta, \qquad (2)$$

where s_N^* is defined as follows. Consider:

(i) *the δ_i's for which $f(x_N^*, \delta_i) > 0$ (violated scenario constraints)*;

(ii) and add to them *additional δ_i's so that a program like* (1) *with only the constraints in (i) and (ii) in place returns the same x_N^* as the original program* (1).

Then, s_N^* is the total number of δ_i's in (i) and in (ii). ⋆

Proof: See Section III-A. ■

To apply Theorem 1, one does not need to know probability $\mathbb{P}$: one solves (1) and finds s_N^*, this is plugged into function $\epsilon(\cdot)$ and the result provides an upper-bound to the violation that is guaranteed with (high) probability $1-\beta$ regardless of what $\mathbb{P}$ is. This may sound somehow magic and one can argue the following: given $(\delta_1, \dots, \delta_N)$, I can always make up a $\mathbb{P}$ that is able to generate this sample and such that relation $V(x_N^*) > \epsilon(s_N^*)$ holds, so how can (2) assert that this has a probability as low as β to happen? The answer the theorem gives is that, no matter how much effort we put in the construction of $\mathbb{P}$, granted that $\mathbb{P}$ can very well give $V(x_N^*) > \epsilon(s_N^*)$ for the sample at hand, still this same $\mathbb{P}$ will be able to generate other samples for which $V(x_N^*) > \epsilon(s_N^*)$ only with a probability that never exceeds β. This is the ground on which one finds protection against an excess of violation without using any prior knowledge on $\mathbb{P}$ (*agnostic setup*). Typically β is set to a very small value, like $\beta = 10^{-7}$, so that it is practically certain that $V(x_N^*) \leq \epsilon(s_N^*)$.[3]

Interestingly, Theorem 1 also offers a rigorous tool to select the hyper-parameter ρ. To this end, one tries out values of ρ in a grid and compares the corresponding solutions in terms of cost $c(x_N^*)$ (which is readily available as an outcome of the optimization problem) and risk of violation (as evaluated through $\epsilon(s_N^*)$) to make a suitable selection.[4]

[1] This is the same as a total ordering defined over $\mathcal{X}$. Note that it is enough to break the tie on x because, at optimum, it must be that $f(x, \delta_i) = \xi_i$, which uniquely identify the ξ_i variables once the tie on x is broken.

[2] The fact that the root is unique is easily seen; see, e.g., [14, footnote 3].

[3] Although the use of this bound in the setup of this paper is new, function $\epsilon(\cdot)$ in (2) appeared in previous contributions in other contexts. Moreover, the properties of $\epsilon(\cdot)$ have been extensively studied and the informativeness of the bound $\epsilon(s_N^*)$ has been discussed, e.g., in [15], [12], [13].

[4] In this process, the user has to pay attention to the fact that each single evaluation of the risk may fail to be correct with probability β; hence, all evaluations, and thereby the evaluation for the selection that has been made, are simultaneously guaranteed with confidence $1 - M\beta$, where M is the total number of evaluations. This is not a big concern since enforcing very low values of β impacts quite marginally on $\epsilon(k)$.

Remark 1 (On the computation of s_N^*): Although s_N^* is an observable that can be computed from $\delta_1, \dots, \delta_N$, one may rightly notice that the computation of s_N^* according to its definition can be extremely demanding.[5] Nonetheless, s_N^* can be over-bounded with relative low effort. To this end, it is enough to scan progressively, one by one, the scenarios in program (1) for which $\xi_i^* \le 0$, and each time one tries to discard the corresponding constraint. If the x-part of the solution changes, then the scenario is re-instated, otherwise, if the the x-part of the solution keeps the same, the scenario is actually discarded and the procedure continues by considering the next scenario. At the end, one is left with a sub-sample of scenarios whose number upper-bounds s_N^*. Provably, $\epsilon(k)$ in Theorem 1 is a monotonic function of k, so over-bounding s_N^* and using this bound in $\epsilon(k)$ leads to evaluations that, while somehow loose, are still statistically valid. ⋆

B. The risk of exceeding the constraint function levels

$V(x)$ is a primary indicator of the quality of a given decision x, since it indicates the risk of not achieving the threshold of satisfaction when the obtained solution faces a new uncertainty instance. However, given that our starting point was a finer description of the interaction between x and δ as given by $f(x, \delta)$, one may be also interested in characterizing the probability of violating the condition $f(x, \delta) \le \ell$ for values of ℓ other than 0. This leads to the following extended notion of risk:

$$V(x, \ell) := \mathbb{P}\{\delta : \; f(x, \delta) > \ell\},$$

which is the probability of exceeding level ℓ when the decision is x. The previous notion of risk is recovered when $\ell = 0$.

In the context of (1), it is possible to provide assessments – again entirely data-driven – of $V(x_N^*, \ell_N^{*,j})$, where $\ell_N^{*,j}$ is the j-th value (taken in descending order) no greater than 0 achieved by $f(x_m^*, \delta_i)$ for some i. This may provide insightful indications on the *margin* with which constraint $f(x_N^*, \delta) \le 0$ is satisfied. Moreover, following an argument first given in [16, Section 3.3] and also used in [17], this may form the ground by which one can also provide an evaluation of the *cumulative distribution function* of $f(x_N^*, \delta)$ seen as a random variable that depends on δ, so further enriching the assessment of x_N^*.

We start with a theorem and then we briefly discuss how the result in the theorem can be used to evaluate the cumulative distribution function of $f(x_N^*, \delta)$.

Theorem 2: Given a confidence parameter $\beta \in (0, 1)$, for any $k = 0, 1, \dots, N$ let $\epsilon(k)$ be defined as in (1). It holds that

$$\mathbb{P}^N\{V(x_N^*, \ell_N^{*,j}) > \epsilon(s_N^{*,j})\} \le \beta,$$

where $s_N^{*,j}$ is defined as follows. Consider:

(i) *the δ_i's for which $f(x_N^*, \delta_i) > \ell_N^{*,j}$;*
(ii) and add to them *additional δ_i's so that a program like* (1) *with only the constraints in (i) and (ii) in place returns the same x_N^* as the original program* (1).

Then, s_N^* is given by the total number of δ_i's in (i) and in (ii) plus 1 if none of the δ_i's in (ii) is such that $f(x_N^*, \delta_i) = \ell_N^{*,j}$ (the value 1 is not added if this condition is not satisfied). ⋆

Proof: See Section III-B. ■

We next sketch how the cumulative distribution function can be evaluated from the previous result. Start by denoting with $F_x(\ell)$ the cumulative distribution function of $f(x, \delta)$ for a given x; that is, $F_x(\ell) := \mathbb{P}\{f(x, \delta) \le \ell\}$, $\ell \in \mathbb{R}$. Then, consider the result of Theorem 1 and that of Theorem 2 for all the, say $\bar{j}$, levels of $f(x_N^*, \delta_i)$ below zero. By summing up all the confidence parameters, and noticing that $F_x(\ell) = 1 - V(x, \ell)$, one has with confidence $1 - (\bar{j} + 1)\beta$ that $F_{x_N^*}(\ell_N^{*,j}) \ge 1 - \epsilon(s_N^{*,j})$, *simultaneously* for all $j = 0, 1, \dots, \bar{j}$, where we set $\ell_N^{*,0} = 0$ and $s_N^{*,0} = s_N^*$. Since cumulative distribution functions are monotonically increasing, a bound for $F_{x_N^*}(\ell_N^{*,j})$ extends also to values of ℓ greater than $\ell_N^{*,j}$. This yields

$$F_{x_N^*}(\ell) \ge F_\epsilon(\ell), \; \forall \ell,$$

with confidence $1 - (\bar{j} + 1)\beta$, where

$$F_\epsilon(\ell) = \begin{cases} 1 - \epsilon(s_N^*), & \ell \ge 0 \\ 1 - \epsilon(s_N^{*,j}), & \ell^{*,j} \le \ell < \ell^{*,j-1}, \; {\scriptstyle j=1,\dots,\bar{j}} \\ 0, & \ell < \ell^{*,\bar{j}}. \end{cases}$$

This shows that $F_\epsilon(\ell)$ is a, valid with confidence $1-(\bar{j}+1)\beta$, lower bound to the *whole* cumulative distribution function of $f(x_N^*, \delta)$ with respect to the variability of δ.

III. Proofs

A. Proof of Theorem 1

The proof is carried out by invoking Theorem 1 of [14], which is concerned with the characterization of the risk of a decision z_N^* made according to a generic scenario-based decision scheme satisfying certain assumptions. Thus, we have first to frame the setup of the present paper into that of [14]. A convenient formalization amounts to consider as decision z_N^* the value of x_N^* augmented with the number of variables $\xi_{N,i}^*$ that are positive (considering the actual value of $\xi_{N,i}^*$ is redundant for the goal we pursue here). To be precise, let $\mathcal{Z} = \mathcal{X} \times \mathbb{N}$, with $\mathbb{N} = \{0, 1, \dots\}$, be the space hosting the decisions and define $z_N^* = (x_N^*, q_N^*)$, where $q_N^* := \#[\xi_{N,i}^* > 0, \;\; i = 1, \dots, N]$. The map from $\delta_1, \dots, \delta_N$ to z_N^* is indicated with the symbol $M_N^{\rm ocr}$, i.e., $z_N^* = M_N^{\rm ocr}(\delta_1, \dots, \delta_N)$ (superscript ocr stands for *optimization with constraint relaxation*). Further, let $\mathcal{Z}_\delta := \{(x, q) \in \mathcal{Z} : \; f(x, \delta) \le 0\}$ and define $V(z) := \mathbb{P}\{\delta : z \notin \mathcal{Z}_\delta\}$. In view of the definition of $\mathcal{Z}_\delta$ it holds that $V(z) = V(x) = \mathbb{P}\{\delta : \; f(x, \delta) > 0\}$. Thus, by applying the theory of [14], we shall upper bound $V(z_N^*)$, which is the same as $V(x_N^*)$.

[5] A "brute force" approach to compute s_N^* requires to repeatedly solve a problem like (1) over all combinations of constraints that is obtained by always keeping those violated and including or leaving out in all possible ways the others.

Towards this goal, we have to verify that the assumptions of Theorem 1 of [14] are satisfied, a fact that is proven by the following lemma.

Lemma 1: The family of maps $M_N^{\rm ocr}$, $N = 0, 1, \ldots$ satisfies Assumption 1 of [14]. Specifically, for every non-negative integers N and m, and for every choice of $\delta_1, \ldots, \delta_N$, and $\delta_{N+1}, \ldots, \delta_{N+m}$, the following three properties hold:

(i) if $\delta_{i_1}, \ldots, \delta_{i_N}$ is a permutation of $\delta_1, \ldots, \delta_N$, then it holds that $M_N^{\rm ocr}(\delta_1, \ldots, \delta_N) = M_N^{\rm ocr}(\delta_{i_1}, \ldots, \delta_{i_N})$;

(ii) if $z_N^* \in \mathcal{Z}_{\delta_{N+i}}$ for all $i = 1, \ldots, m$, then it holds that $z_{N+m}^* = M_{N+m}^{\rm ocr}(\delta_1, \ldots, \delta_{N+m}) = M_N^{\rm ocr}(\delta_1, \ldots, \delta_N) = z_N^*$;

(iii) if $z_N^* \notin \mathcal{Z}_{\delta_{N+i}}$ for one or more $i = 1, \ldots, m$, then it holds that $z_{N+m}^* = M_{N+m}^{\rm ocr}(\delta_1, \ldots, \delta_{N+m}) \neq M_N^{\rm ocr}(\delta_1, \ldots, \delta_N) = z_N^*$. ⋆

Proof: Condition (i) follows from the fact that x_N^* and q_N^* in the definition of z_N^* do not depend on the ordering of the scenarios.

To show (ii), note that $z_N^* \in \mathcal{Z}_{\delta_{N+i}}$ for all $i = 1, \ldots, m$ means that $f(x_N^*, \delta_{N+i}) \leq 0$ for all $i = 1, \ldots, m$. Consider problem (1) with $N{+}m$ in place of N. Since $f(x_N^*, \delta_{N+i}) \leq 0$ for all $i = 1, \ldots, m$, augmenting the solution of (1) with $\xi_i = 0$, $i = N+1, \ldots, N+m$, gives a point $(x_N^*, \xi_{N,1}^*, \ldots, \xi_{N,N}^*, 0, \ldots, 0)$ that is feasible for problem (1) with $N + m$ in place of N. It is claimed that this is indeed the optimal solution. As a matter of fact, if the optimal solution were a different one, say $(\bar{x}, \bar{\xi}_i, i = 1, \ldots, N{+}m)$, then one of the following two cases would hold:

(a) $c(\bar{x}) + \rho \sum_{i=1}^{N+m} \bar{\xi}_i < c(x_N^*) + \rho \sum_{i=1}^{N} \xi_{N,i}^*$. But then this would give $c(\bar{x}) + \rho \sum_{i=1}^{N} \bar{\xi}_i < c(x_N^*) + \rho \sum_{i=1}^{N} \xi_{N,i}^*$ (because the dropped $\bar{\xi}_i$, $i = N{+}1, \ldots, N{+}m$, are non-negative), showing that in problem (1) $(\bar{x}, \bar{\xi}_i, i = 1, \ldots, N)$ would outperform the optimal solution $(x_N^*, \xi_{N,i}^*, i = 1, \ldots, N)$, which is impossible;

(b) $c(\bar{x}) + \rho \sum_{i=1}^{N+m} \bar{\xi}_i = c(x_N^*) + \rho \sum_{i=1}^{N} \xi_{N,i}^*$ and $\bar{x}$ ranks better than x_N^* according to the tie-break rule used to single out the solution. But then $(\bar{x}, \bar{\xi}_i, i = 1, \ldots, N)$ would be feasible for (1) and would achieve $c(\bar{x}) + \rho \sum_{i=1}^{N} \bar{\xi}_i \leq c(x_N^*) + \rho \sum_{i=1}^{N} \xi_{N,i}^*$. Should this latter equation hold with inequality, we would have a contradiction similarly to (a). If instead equality holds, then $(\bar{x}, \bar{\xi}_i, i = 1, \ldots, N)$ would still be preferred to $(x_N^*, \xi_{N,i}^*, i = 1, \ldots, N)$ in problem (1) because $\bar{x}$ ranks better than x_N^*, leading again to a contradiction.

Therefore, it remains proven that $x_{N+m}^* = x_N^*$, $\xi_{N+m,i}^* = \xi_{N,i}^*$ for $i = 1, \ldots, N$ and $\xi_{N+m,i}^* = 0$ for $i = N + 1, \ldots, N + m$. This gives $z_{N+m}^* = (x_{N+m}^*, q_{N+m}^*) = (x_N^*, q_N^*) = z_N^*$, which shows the validity of (ii).

Consider now (iii) and suppose instead that $z_N^* \notin \mathcal{Z}_{\delta_{N+i}}$ for some i, i.e., $f(x_N^*, \delta_{N+i}) > 0$ for some i. Then, if it happens that $x_{N+m}^* = x_N^*$, then $\xi_{N+m,i}^* = \xi_{N,i}^*$ for $i = 1, \ldots, N$ and $\xi_{N+m,N+i}^* > 0$ for some i. Whence, $q_{N+m}^* > q_N^*$, which implies that $z_{N+m}^* \neq z_N^*$. If instead $x_{N+m}^* \neq x_N^*$, this gives straightforwardly $z_{N+m}^* \neq z_N^*$. This proves the validity of (iii). ■

Since Assumption 1 of [14] holds true for $M_N^{\rm ocr}$, Theorem 1 of [14] can now be invoked to claim that

$$\mathbb{P}^N\{V(x_N^*) > \epsilon(\tilde{s}_N^*)\} = \mathbb{P}^N\{V(z_N^*) > \epsilon(\tilde{s}_N^*)\} \leq \beta,$$

where $\tilde{s}_N^*$ is the so-called complexity of $M_N^{\rm ocr}(\delta_1, \ldots, \delta_N)$, i.e., it is the cardinality of a minimal sub-sample $\delta_{i_1}, \ldots, \delta i_k$ of $\delta_1, \ldots, \delta_N$ such that $M_k^{\rm ocr}(\delta_{i_1}, \ldots, \delta_{i_k}) = M_N^{\rm ocr}(\delta_1, \ldots, \delta_N) = z_N^*$. To conclude the proof notice that all the δ_i's for which $f(x_N^*, \delta_i) > 0$ (corresponding to $\xi_{N,i}^* > 0$) must be part of the $\delta_{i_1}, \ldots, \delta i_k$ above. Indeed, if not, at x_N^* there would be a deficiency of violated constraints so that, even though $M_k(\delta_{i_1}, \ldots, \delta_{i_k})$ would return x_N^*, a value of q strictly lower than q_N^* would be achieved, yielding $M_k(\delta_{i_1}, \ldots, \delta_{i_k}) \neq z_N^*$. Thus, $\delta_{i_1}, \ldots, \delta_{i_k}$ must contain all the δ_i's for which $f(x_m^*, \delta_i) > 0$. Once this is recognized, it is then apparent that in order to retrieve z_N^* with $\delta_{i_1}, \ldots, \delta_{i_k}$ it is enough to secure that $M_k(\delta_{i_1}, \ldots, \delta_{i_k})$ return x_N^*. Hence, we have to add to the δ_i's for which $f(x_m^*, \delta_i) > 0$ a minimal amount of other δ_i's such that solving (1) with only the scenarios $\delta_{i_1}, \ldots, \delta_{i_k}$ in place gives x_N^* as x component of the solution. This shows that $\tilde{s}_N^* = s_N^*$, so yielding $\mathbb{P}^N\{V(x_N^*) > \epsilon(s_N^*)\} \leq \beta$. □

B. Proof of Theorem 2

We want again to rely on Theorem 1 of [14]. Since the departure from the proof of Theorem 1 is minor, to avoid repetitions, only the main ideas are provided and some details are left to the reader.

This time, let $\mathcal{Z} = \mathcal{X} \times \mathbb{R} \times \mathbb{N}$ and define $z_N^{*,j} = (x_N^*, \ell_m^{*,j}, q_N^{*,j})$, where $q_N^{*,j}$ is the number of indexes $i \in \{1, \ldots, N\}$ for which $f(x_N^*, \delta_i) > \ell_N^{*,j}$. The map from $\delta_1, \ldots, \delta_N$ to $z_N^{*,j}$ is indicated with the symbol $M_N^{{\rm ocr},j}$. Let $\mathcal{Z}_\delta := \{(x, \ell, q) \notin \mathcal{Z} : f(x, \delta) \leq \ell\}$ and so that $V(z) := \mathbb{P}\{\delta : z \in \mathcal{Z}_\delta\}$ is such that $V(z) = V(x, \ell) = \mathbb{P}\{\delta : f(x, \delta) > \ell\}$.

Mutatis mutandis, an argument conceptually identical to that used to prove Lemma 1 allows one to show that the three properties (i)-(iii) in Lemma 1 hold true for $M_N^{{\rm ocr},j}$. Thus, the family of maps $M_N^{{\rm ocr},j}$ satisfy the assumptions for Theorem 1 of [14], which can now be invoked to claim that

$$\mathbb{P}^N\{V(x_N^*, \ell_N^{*,j}) > \epsilon(\tilde{s}_N^{*,k})\} = \mathbb{P}^N\{V(z_N^{*,k}) > \epsilon(\tilde{s}_N^{*,k})\} \leq \beta,$$

where $\tilde{s}_N^{*,j}$ is the complexity of $M_N^{{\rm ocr},j}(\delta_1, \ldots, \delta_N)$ (i.e., the cardinality of a minimal sub-sample $\delta_{i_1}, \ldots, \delta i_k$ of $\delta_1, \ldots, \delta_N$ such that $M_k^{{\rm ocr},j}(\delta_{i_1}, \ldots, \delta_{i_k}) = M_N^{{\rm ocr},j}(\delta_1, \ldots, \delta_N)$).

To conclude the proof, note that all the δ_i's such that $f(x_N^*, \delta_i) > \ell_N^{*,j}$ must be included in the $\delta_{i_1}, \ldots, \delta i_k$, because otherwise, even assuming that $M_k^{{\rm ocr},j}(\delta_{i_1}, \ldots, \delta_{i_k})$ returns x_N^* as x component of the decision, we would not have enough δ_i's to reconstruct $q_N^{*,j}$. Once all the δ_i's such that $f(x_N^*, \delta_i) > \ell_N^{*,j}$ are taken, to complete $\delta_{i_1}, \ldots, \delta i_k$, we need to introduce a minimal number of δ_i's from the remaining ones such that solving (1) with only the chosen scenarios in place gives x_N^* as x component of the solution.

Moreover, if not already present in the selected scenarios, we have to add one δ_i such that $f(x_N^*, \delta_i) = \ell_N^{*,j}$ in order to be able to reconstruct $\ell_N^{*,j}$ from the $\delta_{i_1}, \dots, \delta i_k$.

Thus, altogether, we have that $\tilde{s}_N^{*,j} = s_N^{*,j}$, so showing that $\mathbb{P}^N\{V(x_N^*, \ell_N^{*,j}) > \epsilon(s_N^{*,j})\} \leq \beta$. □

IV. Simulation Example

We consider a slightly modified version of the mixed-integer, finite-horizon, open-loop input design problem discussed in [18]. As observed in [18], problems of this type are common as single step of MPC design schemes. They also arise in sensor-less environments where no feedback is available. The example we present here is just a toy version of these problems that has the purpose of illustrating the theory.

Consider the discrete-time uncertain linear system

$$\eta(t+1) = A\eta(t) + Bu(t), \qquad \eta(0) = \eta_0, \tag{3}$$

where $\eta(t) \in \mathbb{R}^2$ is the state variable, $u(t) \in \mathbb{R}$ is the control input, $\eta_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$, $B = \begin{bmatrix} 0 & 0.25 \end{bmatrix}^\top$, and $A \in \mathbb{R}^{2\times 2}$ is an uncertain state matrix (in this example, we identify δ with A). The entries of A are generated as draws of four independent Gaussian distributions with means

$$\bar{A} = \begin{bmatrix} 0.8 & -1 \\ 0 & -0.9 \end{bmatrix},$$

and standard deviations equal to $0.025(1+v_i)$, $i = 1, 2, 3, 4$, where v_i are draws from four independent χ_1^2 distributions. Notice that the distribution according to which A is generated is given for reproducibility purposes only, but in no way it is used to design the solution or for quality evaluation. The only information we rely on is given by $N = 1000$ realizations of A, that is, $A_1, \dots, A_{1000}$, which are the scenarios. Also, due to actuation constraints, the input $u(t)$ must be chosen from a *finite* set $\mathcal{U} := \{-10, \cdots, -1, 0, 1, \cdots, 10\}$, a fact that leads to a mixed-integer setup.

Informally, the control objective is to choose the input sequence $u(0), \cdots, u(T-1)$ so as to drive the system state at time $T = 5$ as close as possible to the origin, where the distance is measured according to the maximum norm $\|\eta(T)\|_\infty := \max(|\eta_1(T)|, |\eta_2(T)|)$. Since $\eta(T) = A^T\eta_0 + \sum_{t=0}^{T-1} A^{T-1-t}Bu(t)$, if we let $R = \begin{bmatrix} B & AB & \cdots & A^{T-1}B \end{bmatrix}$ and $\boldsymbol{u} = \begin{bmatrix} u(T-1) & u(T-2) & \cdots & u(0) \end{bmatrix}^\top$, we have that $\|\eta(T)\|_\infty = \|A^T\eta_0 + R\boldsymbol{u}\|_\infty$. The input design problem is then formulated as

$$\min_{\substack{h\geq 0, \boldsymbol{u}\in\mathcal{U}^T,\\ \xi_i \geq 0}} \quad h + \rho \sum_{i=1}^N \xi_i \tag{4}$$

$$\text{subject to:} \quad \|A_i^T\eta_0 + R_i\boldsymbol{u}\|_\infty - h \leq \xi_i, \quad i = 1, \dots, N,$$

which is a specific instance of (1). The interpretation of (4) is as follows. Consider first a modified problem that has no ξ_i's. Then, h is an upper bound to $\|A_i^T\eta_0 + R_i\boldsymbol{u}\|_\infty$, $i = 1, \dots, N$ and, hence, (4) performs a minimization that is robust over the scenarios. By relaxing the constraints with the ξ_i's, some realizations are allowed to exceed h, which is expected to be beneficial to tone down the importance of nasty scenarios, so yielding a better threshold h and $\eta(T)$ closer to the origin for all the other scenarios.

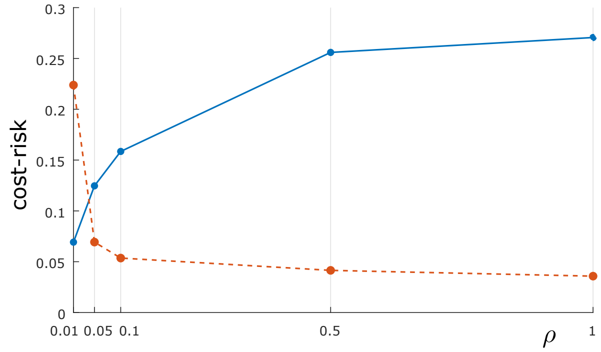

Fig. 1. The cost-risk plot: h_N^* (solid line) vs. $\epsilon(s_N^*)$ (dashed line) as functions of ρ.

Problem (4) was solved for various values of ρ in the grid $\{0.01, 0.05, 0.1, 0.5, 1\}$, each time obtaining a distinct solution $\boldsymbol{u}_N^*, h_N^*$. Clearly, a sensible choice of one of these solutions cannot be only based on h_N^*, this value has to be paired with the risk $V(\boldsymbol{u}_N^*, h_N^*)$, that is, the probability that $\|\eta(T)\|_\infty$ exceeds h_N^* when using $\boldsymbol{u}_N^*$ (in the present context, we have that $V(\boldsymbol{u}, h) = \mathbb{P}\{A : \|A^T\eta_0 + R\boldsymbol{u}\|_\infty - h > 0\}$). To estimate this probability, when solving (4), we also evaluated s_N^* (following the indications in Remark 1) and, according to Theorem 1, we upper-bounded $V(\boldsymbol{u}_N^*, h_N^*)$ via $\epsilon(s_N^*)$ (β was set to the value 10^{-7}). The obtained values of h_N^* and $\epsilon(s_N^*)$, along with s_N^* and the number of the $\xi_{i,N}^*$ bigger than zero are reported in the following table:

ρ	h_N^*	$\epsilon(s_N^*)$	s_N^*	$\#[\xi_{i,N}^* > 0]$
0.01	0.07	22.38 %	148	99
0.05	0.12	6.93 %	27	19
0.1	0.16	5.36 %	17	9
0.5	0.26	4.15 %	10	1
1	0.27	3.58 %	7	0

(notice that $\#[\xi_{i,N}^* > 0] = 0$ for $\rho = 1$, i.e. the robust-over-the-scenarios solution is obtained in this case). The values of h_N^* and $\epsilon(s_N^*)$ for various ρ are also graphically displayed in Figure 1 (we call this a cost-risk plot). Figure 1 provides information on the various alternatives and one can make a choice correspondingly, depending also on one's attitude towards risk. For example, one may want to pick $\rho = 0.1$ because the cost has become small enough (it is half the cost for $\rho = 1$) and the risk is still moderate.

To obtain a better evaluation of the various solutions, one may consider the cumulative distribution function of $\|A^T\eta_0 + R\boldsymbol{u}_N^*\|_\infty - h_N^*$ as provided by Theorem 2 and the analysis in Section II-B. Towards this goal, for $\rho = 0.1$, $\rho = 0.5$, and $\rho = 1$ we also computed $F_\epsilon(\ell)$ (which with high confidence lower bounds the cumulative distribution function of $\|A^T\eta_0 + R\boldsymbol{u}_N^*\|_\infty - h_N^*$). For easier interpretation, Figure 2 profiles in solid line $F_\epsilon(\ell' - h_N^*)$ for the three cases, these are shifted versions of the corresponding $F_\epsilon(\ell)$'s and provide lower bounds to the cumulative distribution function

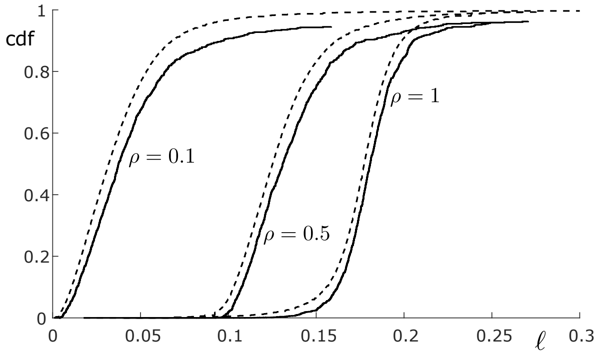

Fig. 2. $F_\epsilon(\ell' - h_N^*)$ (solid lines) vs. actual cumulative distribution function of $\|\eta(T)\|_\infty$ (dashed lines) for $\rho = 0.1$, $\rho = 0.5$, and $\rho = 1$.

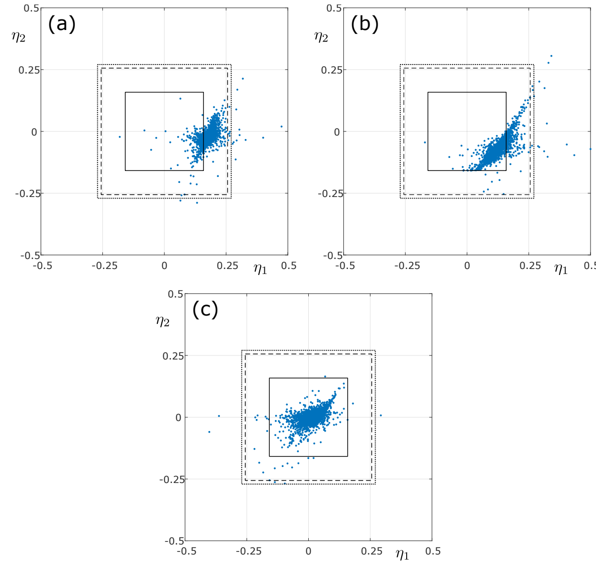

Fig. 3. Realizations of $\eta(T)$ for the solution corresponding to: (a) $\rho = 1$, (b) $\rho = 0.5$, (c) $\rho = 0.1$. In all panels we have represented the boxes $\|\eta\|_\infty \leq h_N^*$ for $\rho = 1$ (dotted line), $\rho = 0.5$ (dashed line), and $\rho = 0.1$ (solid line).

of $\left\|A^T\eta_0 + R\boldsymbol{u}_N^*\right\|_\infty = \|\eta(T)\|_\infty$. From an analysis of this figure, one can obtain insightful information. For instance, using the table on the previous page one might be in doubt as to weather $\rho = 0.1$ should be preferred to $\rho = 1$; indeed, $\rho = 1$ gives a higher h_N^* but the constructed box contains all data points corresponding to a lower risk of overstepping its boundaries (in Figure 2, this fact corresponds to having a higher upper bound in $\ell' = 0.27$ for $\rho = 1$ than that in $\ell' = 0.16$ for $\rho = 0.1$). While this is true, an inspection of Figure 2 also reveals that the solution corresponding to $\rho = 0.1$ yields values of $\|\eta(T)\|_\infty$ that are more concentrated in the low-value range than for $\rho = 0.5$ and $\rho = 1$, from which our preference can lean towards selecting $\rho = 0.1$.

Taking advantage of the fact that this example is *in silico*, we performed a validation of the analysis above. Specifically, 100000 new scenarios A_i were drawn and the actual cumulative distribution function of $\|\eta(T)\|_\infty$ was Monte Carlo evaluated for the solutions corresponding to $\rho = 0.1$, $\rho = 0.5$, and $\rho = 1$. This gave the dashed lines depicted in Figure 2. In all of the three cases, we have that the actual distribution lies above $F_\epsilon(\ell' - h_N^*)$, as predicted by Theorem 2, and is close enough to $F_\epsilon(\ell' - h_N^*)$ to show that the bounds are informative. This latter point can be also appreciated from Figure 3, which depicts the realizations of $\eta(T)$ corresponding to the first 10000 newly extracted scenarios when $\rho = 0.1$, $\rho = 0.5$, and $\rho = 1$.

References

[1] M. Campi and S. Garatti, *Introduction to Scenario Oprimization*, ser. MOS-SIAM series on Optimization. SIAM, 2018.

[2] M. Campi, A. Carè, and S. Garatti, "The scenario approach: a tool at the service of data-driven decision making," *Annual Reviews in Control*, vol. 52, pp. 1–17, 2021.

[3] 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*, vol. 50, no. 12, pp. 3009–3018, 2014.

[4] S. Grammatico, X. Zhang, K. Margellos, P. Goulart, and J. Lygeros, "A scenario approach for non-convex control design," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 334–345, 2016.

[5] P. Mohajerin Esfahani, T. Sutter, and J. Lygeros, "Performance bounds for the scenario approach and an extension to a class of non-convex programs," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 46–58, 2015.

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

[7] L. G. Crespo, S. P. Kenny, and D. P. Giesy, "Interval predictor models with a linear parameter dependency," *Journal of Verification, Validation and Uncertainty Quantification*, vol. 1, no. 2, pp. 1–10, 2016.

[8] M. Assif, D. Chatterjee, and R. Banavar, "Scenario approach for minmax optimization with emphasis on the nonconvex case: Positive results and caveats," *SIAM Journal on Optimization*, vol. 30, no. 2, pp. 1119–1143, 2020.

[9] C. Shang and F. You, "A posteriori probabilistic bounds of convex scenario programs with validation tests," *IEEE Transactions on Automatic Control*, vol. 66, no. 9, pp. 4015–4028, 2021.

[10] A. Falsone, K. Margellos, J. Zizzo, M. Prandini, and S. Garatti, "On the sensitivity of linear resource sharing problems to the arrival of new agents," *IEEE Transactions on Automatic Control*, vol. 68, no. 1, pp. 272–284, 2023.

[11] L. Romao, K. Margellos, and A. Papachristodoulou, "Probabilistic feasibility guarantees for convex scenario programs with an arbitrary number of discarded constraints," *Automatica*, vol. 149, p. 110601, 2023.

[12] S. Garatti and M. Campi, "Risk and complexity in scenario optimization," *Mathematical Programming*, vol. 191, no. 1, pp. 243–279, 2022.

[13] M. Campi and S. Garatti, "A theory of the risk for optimization with relaxation and its application to support vector machines," *Journal of Machine Learning Research*, vol. 22, no. 288, pp. 1–38, 2021.

[14] S. Garatti and M. C. Campi, "The risk of making decisions from data through the lens of the scenario approach," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 607–612, 2021, 19th IFAC Symposium on System Identification SYSID 2021.

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

[16] A. Carè, S. Garatti, and M. Campi, "Scenario min-max optimization and the risk of empirical costs," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2061–2080, 2015.

[17] G. Arici, M. Campi, A. Carè, M. Dalai, and F. Ramponi, "A theory of the risk for empirical CVaR with application to portfolio selection." *Journal of Systems Science and Complexity*, vol. 34, no. 5, pp. 1879–1894, 2021.

[18] M. Campi, S. Garatti, and F. Ramponi, "A general scenario theory for nonconvex optimization and decision making," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4067–4078, 2018.