On the Road Between Robust Optimization and the Scenario Approach for Chance Constrained Optimization Problems

Kostas Margellos, Paul Goulart, and John Lygeros

***Abstract*—We propose a new method for solving chance constrained optimization problems that lies between robust optimization and scenario-based methods. Our method does not require prior knowledge of the underlying probability distribution as in robust optimization methods, nor is it based entirely on randomization as in the scenario approach. It instead involves solving a robust optimization problem with bounded uncertainty, where the uncertainty bounds are randomized and are computed using the scenario approach. To guarantee that the resulting robust problem is solvable we impose certain assumptions on the dependency of the constraint functions with respect to the uncertainty and show that tractability is ensured for a wide class of systems. Our results lead immediately to guidelines under which the proposed methodology or the scenario approach is preferable in terms of providing less conservative guarantees or reducing the computational cost.**

***Index Terms*—Chance constrained optimization, randomized algorithms, robust optimization, scenario approach.**

I. Introduction

Robust optimization has attracted increasing attention due to its ability to offer performance guarantees for optimization problems in the presence of uncertainty. Robust control design requires the construction of a decision such that the constraints are satisfied for all admissible values of some uncertain parameter. For such problems, [1]–[4] provide conditions under which the robust variants of standard programming problems are tractable.

An alternative approach is to interpret robustness in a probabilistic sense, allowing for constraint violation with a low probability. This gives rise to chance-constrained optimization problems [5], that, aside from a few special cases [6], are computationally intractable since they require the computation of multi-dimensional probability integrals. To overcome this difficulty, [3], [7] follow a different approach; a robust problem with bounded uncertainty is solved, where the uncertainty bounds are chosen based on certain assumptions on the probability distribution.

Randomization of uncertainty offers an alternative way to provide probabilistic performance guarantees, without assumptions on the probability distribution (see [8]–[14] and references therein). Typically it involves sampling the uncertainty and substituting the chance constraint with a finite number of hard constraints, corresponding to the different uncertainty realizations. To provide probabilistic guarantees based on a finite number of samples, [15]–[18] concentrate on problems that are convex with respect to the decision variables and introduce the so called *scenario approach*.

Manuscript received March 05, 2013; revised October 16, 2013 and January 16, 2014; accepted January 17, 2014. Date of publication January 28, 2014; date of current version July 21, 2014. This work was supported by the European Commission under the project MoVeS, FP7-ICT-257005, the Network of Excellence HYCON2, FP7-ICT-257462, and the Swiss Secretariat for Education and Research under a project within COST-Action IC0806. Recommended by Associate Editor M. Prandini.

The authors are with the Automatic Control Laboratory, Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich 8092, Switzerland (e-mail: margellos@control.ee.ethz.ch; goulart@control.ee.ethz.ch; lygeros@control.ee.ethz.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2014.2303232

The number of samples required to achieve certain probabilistic performance grow linearly with the number of decision variables. In view of reducing the sample size we propose here a hybrid methodology, which does not rely entirely on randomization as in the case of the standard scenario approach, nor does it require knowledge about the uncertainty probability distribution or ad-hoc truncation as in standard robust methods. As in [3], a robust problem with bounded uncertainty is solved, but the uncertainty bounds are computed using the scenario approach. That way we do not require convexity to provide probabilistic guarantees on the constraint satisfaction, but the resulting robust problem needs to be solvable. To guarantee this we impose certain assumptions on the dependency of the constraint functions with respect to the uncertainty and show that tractability is ensured for a wide class of systems. The number of scenarios that must be generated in our case, however, does not depend on the number of decision variables as in the scenario approach, but rather on the dimension of the uncertainty vector or the number of constraints. This fact leads to guidelines under which each of the methods, when applicable, is preferable in terms of providing less conservative guarantees or reducing the computational cost. We also investigate the performance of our approach against the so called sampling-and-discarding approach [18], [19].

In Section II we recall the standard scenario approach, whereas in Sections III and IV we introduce our alternative methods. Section V compares the proposed approaches with the scenario approach and discusses different alternatives, whereas Section VI summarizes our results.

II. Problem Description

Consider the chance constrained optimization problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^{n_x}} \quad J(x) \\ &\text{subject to}: \quad \mathbb{P}\left(\delta\in\Delta \mid \max_{j=1,\dots,n_m} g_j(x,\delta)\le 0\right)\ge 1-\epsilon \qquad (\mathcal{P}_1)\end{aligned}$$

where $\delta\in\Delta\subseteq\mathbb{R}^{n_\delta}$, $J:\mathbb{R}^{n_x}\to\mathbb{R}$, and $g_j:\mathbb{R}^{n_x}\times\Delta\to\mathbb{R}$ for all $j=1,\dots,n_m$. Any x satisfying the chance constraint of $\mathcal{P}_1$ is referred to as an ϵ-level feasible solution. It is assumed that Δ is endowed with a σ-algebra $\mathcal{D}$, that $\mathbb{P}$ is a probability measure defined over $\mathcal{D}$, and that for all $x\in\mathbb{R}^{n_x}$, every $g_j(x,\cdot)$ is measurable with respect to $\mathcal{D}$ and the Borel σ-algebra over $\mathbb{R}$.

The standard scenario approach [15] substitutes the chance constraint in $\mathcal{P}_1$ with a finite number of hard constraints, each corresponding to a different realization $\delta^{(k)}$, $k=1,\dots,N$ of the uncertain parameter δ, extracted according to $\mathbb{P}$. This leads to

$$\begin{aligned} &\min_{x\in\mathbb{R}^{n_x}} \quad J(x) \\ &\text{subject to}: \quad \max_{j=1,\dots,n_m} g_j\left(x,\delta^{(k)}\right)\le 0, \text{ for } k=1,\dots,N. \qquad (\mathcal{P}_1')\end{aligned}$$

Assumption 1: The optimization problem $\mathcal{P}_1'$ is feasible for all possible multi-sample extractions $(\delta^{(1)},\dots,\delta^{(N)})\in\Delta^N$ and its feasibility region has a non-empty interior. Moreover, the solution x^* of $\mathcal{P}_1'$ exists and is unique.

Note that both the uniqueness [15] and feasibility [18] conditions can be relaxed, so Assumption 1 serves only to streamline the presentation of our results. Under Assumption 1, for a given violation level $\epsilon\in(0,1)$ and confidence $\beta\in(0,1)$, select N according to

$$\sum_{k=0}^{n-1}\binom{N}{k}\epsilon^k(1-\epsilon)^{N-k}\le\beta \qquad (1)$$

with $n = n_x$. Equation (1) requires that the tail of a binomial distribution is bounded by the desired β, and is tight for the class of fully supported problems.

Theorem 1. ([16, Th. 1]): Under Assumption 1, if $J(\cdot)$ is convex and $g_j(\cdot, \delta)$, $j = 1, \ldots, n_m$, is convex for every $\delta \in \Delta$, and N is selected according to (1) with $n = n_x$, then the optimal solution x^* of $\mathcal{P}_1'$ is an ϵ-level feasible solution for $\mathcal{P}_1$ with probability at least $1 - \beta$.

Assumption 1 is trivially satisfied in the cases in which we will invoke Theorem 1 below. If $\widetilde{V}(x) = \mathbb{P}(\delta \in \Delta | \max_{j=1,\ldots,n_m} g_j(x, \delta) > 0)$ denotes the probability of constraint violation, then Theorem 1 implies that $\mathbb{P}^N((\delta^{(1)}, \ldots, \delta^{(N)}) \in \Delta^N | \widetilde{V}(x^*) \le \epsilon) \ge 1 - \beta$, where $\mathbb{P}^N$ is the product probability measure.

III. Method 1: Unstructured Constraints

A. Formulation

The main idea of our work is to focus first only on the uncertainty and solve a random program that returns a set B^* that, with certain confidence, encloses a predefined portion of the probability mass of the uncertainty. We then solve the robust counterpart of $\mathcal{P}_1$ where the uncertainty δ is now confined in B^*. We will construct B^* as a hyper-rectangle with outward normals aligned with the canonical basis vectors in $\mathbb{R}^{n_\delta}$. To this end, define constants $\epsilon_i \in (0, 1)$ for $i = 1, \ldots, n_\delta$, such that $\sum_{i=1}^{n_\delta} \epsilon_i = \epsilon$. We seek element-wise bounds $\tau_i := (\underline{\tau}_i, \overline{\tau}_i) \in \mathbb{R}^2$ such that $\delta_i \in [\underline{\tau}_i, \overline{\tau}_i]$ with probability at least $1 - \epsilon_i$, where $\delta_i \in \mathbb{R}$ denotes the ith element of the uncertainty vector δ. Therefore, we consider the family of problems

$$\begin{aligned} &\min_{\tau_i \in \mathbb{R}^2} && \left(\overline{\tau}_i - \underline{\tau}_i\right) \\ &\text{subject to}: && \mathbb{P}\left(\delta \in \Delta \mid \delta_i \in [\underline{\tau}_i, \overline{\tau}_i]\right) \ge 1 - \epsilon_i. \end{aligned} \tag{$\mathcal{P}_2$}$$

The problems in $\mathcal{P}_2$ trivially satisfy Assumption 1 (in particular they are fully supported) and both their objective function and the constraints are convex with respect to the decision variables. Therefore, we can construct a solution using Theorem 1. Since $\mathcal{P}_2$ has only two decision variables, choose N_i from (1) with $n = 2$ and consider the problems

$$\begin{aligned} &\min_{\tau_i \in \mathbb{R}^2} && (\overline{\tau}_i - \underline{\tau}_i) \\ &\text{subject to}: && \delta_i^{(k)} \in [\underline{\tau}_i, \overline{\tau}_i], \text{ for } k = 1, \ldots, N_i \end{aligned} \tag{$\mathcal{P}_2'$}$$

where $\delta_i^{(k)}$ denotes the element i of the sample k. In total $N = \max_{i=1,\ldots,n_\delta} N_i$ samples must be extracted; for each such problem we then choose arbitrarily a subset of these samples with cardinality N_i. For $i = 1, \ldots, n_\delta$, $\tau_i^* := (\underline{\tau}_i^*, \overline{\tau}_i^*)$ is a feasible solution for $\mathcal{P}_2$ with probability at least $1 - \beta_i$. This implies that $\mathbb{P}^{N_i}((\delta^{(1)}, \ldots, \delta^{(N_i)}) \in \Delta^{N_i} | V(\tau_i^*) \le \epsilon_i) \ge 1 - \beta_i$, where $V(\tau_i) = \mathbb{P}(\delta \in \Delta | \delta_i \notin [\underline{\tau}_i, \overline{\tau}_i])$ is the probability of constraint violation. Construct now the hyper-rectangle $B^* := \times_{i=1}^{n_\delta} [\underline{\tau}_i^*, \overline{\tau}_i^*]$ and pose the following robust version of $\mathcal{P}_1$:

$$\begin{aligned} &\min_{x \in \mathbb{R}^{n_x}} && J(x) \\ &\text{subject to}: && \max_{j=1,\ldots,n_m} \max_{\delta \in B^* \cap \Delta} g_j(x, \delta) \le 0. \end{aligned} \tag{$\mathcal{P}_3$}$$

Note that we only need to solve $\mathcal{P}_2'$ once and we can then use its solution B^* for any robust problem $\mathcal{P}_3$, inheriting the same probabilistic performance guarantees. One could alternatively construct a set B^* using some other representation requiring fewer or more decision variables in $\mathcal{P}_2$, e.g., a spherical or ellipsoidal cover for the extracted scenarios. The number of variables required to parameterize any particular geometric representation for B^* dictates the number of scenarios in $\mathcal{P}_2'$. Our proofs can be extended to such cases.

Proposition 1: Suppose that ϵ, $\beta \in (0, 1)$ and ϵ_i, $\beta_i \in (0, 1)$, $i = 1, \ldots, n_\delta$, are chosen such that $\epsilon = \sum_{i=1}^{n_\delta} \epsilon_i$, $\beta = \sum_{i=1}^{n_\delta} \beta_i$, and N_i is chosen according to (1) with $n = 2$. If x^* is a feasible solution of $\mathcal{P}_3$, then x^* is also an ϵ-level feasible solution of $\mathcal{P}_1$, with probability at least $1 - \beta$.

Proof: It suffices to show that for $N = \max_{i=1,\ldots,n_\delta} N_i$, $\mathbb{P}^N((\delta^{(1)}, \ldots, \delta^{(N)}) \in \Delta^N | \widetilde{V}(x^*) \le \epsilon, \text{for all } x^* \in \mathcal{X}_N) \ge 1 - \beta$, where $\mathcal{X}_N$ is the feasibility region of $\mathcal{P}_3$ (it depends on the multi-sample via B^*). If $x^* \in \mathcal{X}_N$ is a feasible solution of $\mathcal{P}_3$ then it will satisfy its constraints, so $\max_{j=1,\ldots,n_m} \max_{\delta \in B^* \cap \Delta} g_j(x^*, \delta) \le 0$. By interchanging the two max operators, we have that if $\delta \in B^* \cap \Delta$ then $\max_{j=1,\ldots,n_m} g_j(x^*, \delta) \le 0$. Hence

$$\begin{aligned} 1 - \widetilde{V}(x^*) &= \mathbb{P}\Big(\delta \in \Delta \mid \max_{j=1,\ldots,n_m} g_j\Big(x^*, \delta\Big) \le 0\Big), \\ &\ge \mathbb{P}\Big(\delta \in \Delta \mid \delta \in B^*\Big) \\ &= 1 - \mathbb{P}\Big(\bigcup_{i=1}^{n_\delta} \Big(\delta \in \Delta \mid \delta_i \notin [\underline{\tau}_i^*, \overline{\tau}_i^*]\Big)\Big), \\ &\ge 1 - \sum_{i=1}^{n_\delta} \mathbb{P}\Big(\delta \in \Delta \mid \delta_i \notin [\underline{\tau}_i^*, \overline{\tau}_i^*]\Big). \end{aligned} \tag{2}$$

The last statement implies that $\widetilde{V}(x^*) \le \sum_{i=1}^{n_\delta} V(\tau_i^*)$. Since this holds for all $x^* \in \mathcal{X}_N$, we have

$$\begin{aligned} &\mathbb{P}^N\Big(\Big(\delta^{(1)}, \ldots, \delta^{(N)}\Big) \in \Delta^N \mid \widetilde{V}\Big(x^*\Big) \le \epsilon, \text{ for all } x^* \in \mathcal{X}_N\Big), \\ &\ge \mathbb{P}^N\Big(\Big(\delta^{(1)}, \ldots, \delta^{(N)}\Big) \in \Delta^N \mid \sum_{i=1}^{n_\delta} V\Big(\tau_i^*\Big) \le \epsilon\Big), \\ &\ge \mathbb{P}^N\Big(\Big(\delta^{(1)}, \ldots, \delta^{(N)}\Big) \in \Delta^N | V\Big(\tau_i^*\Big) \le \epsilon_i, \ \forall i = 1, \ldots, n_\delta\Big), \\ &= 1 - \mathbb{P}^N\Big(\bigcup_{i=1}^{n_\delta}\Big(\Big(\delta^{(1)}, \ldots, \delta^{(N)}\Big) \in \Delta^N \mid V\Big(\tau_i^*\Big) > \epsilon_i\Big)\Big), \\ &\ge 1 - \sum_{i=1}^{n_\delta} \mathbb{P}^{N_i}\Big(\Big(\delta^{(1)}, \ldots, \delta^{(N_i)}\Big) \in \Delta^{N_i} | V\Big(\tau_i^*\Big) > \epsilon_i\Big), \\ &\ge 1 - \beta \end{aligned} \tag{3}$$

where the first inequality is valid due to (2), and the last two follow from the subadditivity of $\mathbb{P}$ and the implications of Theorem 1 for $\mathcal{P}_2'$, respectively. The selection of the first N_i samples in the above procedure was arbitrary, and any subset of $\delta^{(1)}, \ldots, \delta^{(N)}$ with cardinality N_i could have been chosen instead. The interpretation of this derivation is that the probability of all violation probabilities $V(\tau_i^*)$ being simultaneously bounded by the corresponding ϵ_i is at least $1 - \beta$. □

B. Tractability of the Proposed Method

Proposition 1 provides probabilistic guarantees for the probability of constraint violation of *any* feasible solution of $\mathcal{P}_3$, and not only the optimal one as in the scenario approach. The number of samples that we need to generate when Method 1 is adopted depends on the dimension of the uncertainty and not on the number of decision variables as in the conventional scenario approach. Moreover, unlike Theorem 1, we do not require the functions $J(\cdot), g_j(\cdot, \cdot)$, $j = 1, \ldots, n_m$ to be convex with respect to the decision variables. The reason is that the scenario approach is only adopted to solve $\mathcal{P}_2$, which is trivially convex. However, our method requires solving $\mathcal{P}_3$, which is a robust problem with bounded uncertainty. We next consider two cases in which we can solve $\mathcal{P}_3$. For both alternatives we assume that $B^* \cap \Delta$ is "nice"; this is the case for example if $\Delta = \mathbb{R}^{n_\delta}$ or if Δ is a hyper-rectangular set. In the opposite case tractability of our approach is not guaranteed. Note that if Δ is itself a hyper-rectangle, solving directly a robust problem with the uncertainty confined in Δ will generally be more conservative since in this case the hyper-rectangle B^* generated by our approach will always be inscribed in Δ.

Note that even if by means of the following methods the robust constraints can be reformulated so that $\mathcal{P}_3$ is solvable, the possibly non-convex dependency of the objective and the constraint functions on x may lead to a problem which is difficult to solve. We can then use numerical tools for non-convex optimization that may not return a global minimizer. However, since we provide probabilistic guarantees for any feasible solution of $\mathcal{P}_3$, the obtained solution will satisfy the system constraints with certain probability.

1) Vertex Enumeration: We impose the following assumption.

Assumption 2: For all $x \in \mathbb{R}^{n_x}$ and for all $j = 1, \ldots, n_m$, $g_j(x, \delta)$ achieves its maximum with respect to δ at a vertex of B^*.

Problems whose constraint functions are linear, monotone or convex with respect to the uncertainty constitute problem classes that satisfy Assumption 2. Under Assumption 2, it suffices to enforce the constraints of $\mathcal{P}_3$ only for the uncertainty vectors that correspond to the vertices of B^*. Following this vertex enumeration scheme results in a problem with $n_m 2^{n_\delta}$ constraints in total. Improved results have been obtained for robustness problems affected by interval matrix uncertainty [20], [21], however, with a constraint complexity of $O(2^{n_\delta})$. For cases in which the vertex approach leads to a computationally manageable problem, no additional structure on the objective function and the constraints of the initial problem is required.

2) Tractable Reformulation of $\mathcal{P}_3$: To achieve tractability, [1], [3], [22] focus on a specific class of problems that satisfy the following assumption.

Assumption 3: For all $j = 1, \ldots, n_m$, $g_j(x, \delta)$ is convex and homogeneous (i.e., $g_j(x, \alpha\delta) = \alpha g_j(x, \delta)$ for any $\alpha \in \mathbb{R}$) in δ for any fixed $x \in \mathbb{R}^{n_x}$.

Under Assumption 3,[1] it is shown in [3] that using duality techniques the robust counterpart of certain problem classes (linear programs, quadratic constrained quadratic programs, second order cone programs, semi-definite programs) is tractable and in the same class as the original problem, i.e., robust linear programs remain linear programs, etc. Following Theorem 1 of [3], if $\mathcal{P}_3$ is a linear program, this reformulation does not involve any relaxation and requires $n_m(n_\delta + 1)$ decision variables and $n_m(2n_\delta + 1)$ linear constraints in addition to those of $\mathcal{P}_3$. The approach of [3] was extended in [23] to robust mixed-integer problems. Therefore, the proposed approach for the aforementioned types of convex problems, as well as for mixed-integer problems, leads to a problem with constraint complexity $O(n_\delta)$ and with probabilistic guarantees without assumptions on the probability distribution as in [3].

IV. Method 2: Structured Constraints

A. Formulation

We next consider the particular case where the functions g_j are separable in (x, δ):[2]

Assumption 4: For $j = 1, \ldots, n_m$, $g_j(x, \delta) := p_j(x) q_j(\delta)$, where $p_j : \mathbb{R}^{n_x} \to \mathbb{R}$ and $q_j : \Delta \to \mathbb{R}$.

[1] Note that in [3] a concavity assumption is imposed instead since a max-min and not a min-max problem was considered.

[2] The results of this section are easily generalized to the case where $p_j : \mathbb{R}^{n_x} \to \mathbb{R}^\ell$ and $q_j : \Delta \to \mathbb{R}^\ell$, and $g_j(x, \delta) := \langle p_j(x) q_j(\delta) \rangle$, thus allowing for systems that are affine with respect to the uncertainty functions $q_{j\ell}$ (some of the $q_{j\ell}(\cdot)$ could be made trivial). The number of scenarios in this case would depend on the total number of uncertainty functions, i.e. $n_m \ell$.

Similarly to Section III, we construct a hyper-rectangle B_q^* that encloses the image of a collection of samples $\delta^{(k)}$ under the function $q(\delta) = (q_1(\delta), \ldots, q_{n_m}(\delta))$ with a certain probability. The subscript q indicates that B_q^* contains $q(\delta)$ instead of δ. To this end, we consider the problems

$$\begin{aligned} &\min_{\tau_j \in \mathbb{R}^2} \quad (\overline{\tau}_j - \underline{\tau}_j) \\ &\text{subject to:} \quad \mathbb{P}\left(\delta \in \Delta \mid q_j(\delta) \in [\underline{\tau}_j, \overline{\tau}_j]\right) \geq 1 - \epsilon_j \end{aligned} \qquad (\widetilde{\mathcal{P}}_2)$$

Assumption 1 is trivially satisfied by the problems in $\widetilde{\mathcal{P}}_2$, so using Theorem 1 we have

$$\begin{aligned} &\min_{\tau_j \in \mathbb{R}^2} \quad (\overline{\tau}_j - \underline{\tau}_j) \\ &\text{subject to:} \quad q_j\left(\delta^{(k)}\right) \in [\underline{\tau}_j, \overline{\tau}_j], \text{ for } k = 1, \ldots, N_j, \end{aligned} \qquad (\widetilde{P}_2')$$

where N_j is chosen from (1) with $n=2$. In total $N = \max_{j=1,\ldots,n_m} N_j$ samples must be extracted, and for each problem we choose arbitrarily a subset with cardinality N_j. Moreover, $\mathbb{P}^{N_j}((\delta^{(1)}, \ldots, \delta^{(N_j)}) \in \Delta^{N_j} | V(\tau_j^*) \leq \epsilon_j) \geq 1 - \beta_j$, where $\tau_j^* := (\underline{\tau}_j^*, \overline{\tau}_j^*)$ and $V(\tau_j) = \mathbb{P}(\delta \in \Delta | q_j(\delta) \notin [\underline{\tau}_j^*, \overline{\tau}_j^*])$. Construct now the hyper-rectangle $B_q^* := \times_{j=1}^{n_m} [\underline{\tau}_j^*, \overline{\tau}_j^*]$ and pose the following robust version of $\mathcal{P}_1$:

$$\begin{aligned} &\min_{x \in \mathbb{R}^{n_x}} \quad J(x) \\ &\text{subject to:} \quad \max_{j=1,\ldots,n_m} \max_{q(\delta) \in B_q^* \cap q(\Delta)} p_j(x) q_j(\delta) \leq 0. \end{aligned} \qquad (\widetilde{P}_3)$$

Proposition 2: Suppose throughout that Assumption 4 holds. Then:

1) Assume that ϵ, $\beta \in (0,1)$ and ϵ_j, $\beta_j \in (0,1)$, $j = 1, \ldots, n_m$, are chosen such that $\epsilon = \sum_{j=1}^{n_m} \epsilon_j$, $\beta = \sum_{j=1}^{n_m} \beta_j$, and N_j is chosen according to (1) with $n=2$. If x^* is a feasible solution of $\widetilde{\mathcal{P}}_3$, then x^* is also an ϵ-level feasible solution of $\mathcal{P}_1$, with probability at least $1-\beta$.

2) Assume that x^* is an ϵ-level feasible solution of $\mathcal{P}_1$, and select any $\beta \in (0,1)$ and an integer N such that $\epsilon = 1-(1-\beta)^{1/N}$. Select any (ϵ_j, β_j), $j = 1, \ldots, n_m$, such that $N_j \leq N$, and construct the set B_q^* from $\widetilde{\mathcal{P}}_2'$. Then x^* is a feasible solution of $\widetilde{\mathcal{P}}_3$ with probability at least $1-\beta$.

Proof: 1) The proof is similar to that of Proposition 1 and is omitted for brevity.

2) If x^* is a feasible solution of $\mathcal{P}_1$, then $\mathbb{P}(\delta \in \Delta | \max_{j=1,\ldots,n_m} p_j(x^*) q_j(\delta) \leq 0) \geq 1 - \epsilon$. Select any $\beta \in (0,1)$, and an integer N such that $\epsilon \leq 1 - (1-\beta)^{1/N}$. Then, for N independent uncertainty extractions $\delta^{(k)}$, with $k = 1, \ldots, N$, $\mathbb{P}(\delta^{(k)} \in \Delta | \max_{j=1,\ldots,n_m} p_j(x^*) q_j(\delta^{(k)}) \leq 0) \geq (1-\beta)^{1/N}$. Due to independence, for the joint event we have that

$$\mathbb{P}^N\Bigg(\left(\delta^{(1)}, \ldots, \delta^{(N)}\right) \in \Delta^N \mid \max_{j=1,\ldots,n_m} p_j(x^*) q_j\left(\delta^{(k)}\right) \leq 0, \text{ for all } k = 1, \ldots, N \Bigg) \geq 1 - \beta. \quad (4)$$

Select now ϵ_j, β_j, $j = 1, \ldots, n_m$, such that $N_j \leq N$, and solve $\widetilde{\mathcal{P}}_2'$. Let $\underline{\tau}_j^*, \overline{\tau}_i^*$ denote the solution of $\widetilde{\mathcal{P}}_2'$, use it to construct B_q^*, and formulate $\widetilde{\mathcal{P}}_3$. The argument inside the probability of (4) implies that for all $j = 1, \ldots, n_m$ and $k = 1, \ldots, N$, $p_j(x^*) q_j(\delta^{(k)}) \leq 0$. Therefore, it also holds that for all $j = 1, \ldots, n_m$ and $k = 1, \ldots, N_j$, $p_j(x^*) q_j(\delta^{(k)}) \leq 0$; the choice of the first N_j samples is arbitrary, and any subset of $\delta^{(1)}, \ldots, \delta^{(N)}$ with cardinality N_j could have been selected instead. Since the constraint functions are linear with respect to $q_j(\cdot)$ (Assumption 4), $p_j(x^*) q_j(\delta^{(k)}) \leq 0$ for all $k = 1, \ldots, N_j$ implies that $p_j(x^*) q_j(\delta^{(k)}) \leq 0$ for all $q_j(\delta) \in [\underline{\tau}_j^*, \overline{\tau}_i^*]$. Therefore

$$\mathbb{P}^N\Bigg(\left(\delta^{(1)}, \ldots, \delta^{(N)}\right) \in \Delta^N \mid \max_{j=1,\ldots,n_m} \max_{q(\delta) \in B_q^* \cap q(\Delta)} p_j(x^*) q_j(\delta) \leq 0 \Bigg) \geq 1 - \beta. \quad (5)$$

Statement (5) implies that with probability at least $1-\beta$, x^* is a feasible solution for $\widetilde{\mathcal{P}}_3$. □

Note that an argument similar to the second part of Proposition 2 can not be applied for the more general problem described in Section III. The reason is that the step analogous to that from (4) to (5) would no longer be valid, since the fact that for all $j = 1,\ldots,n_m$ and $k = 1,\ldots,N_j$, $g_j(x^*, \delta^{(k)}) \leq 0$, does not necessarily imply that $\max_{\delta\in B^*\cap\Delta} g_j(x^*,\delta) \leq 0$. An analogous statement can still be made if the constraint functions satisfy Assumption 2.

B. Tractability of the Proposed Method

The first part of Proposition 2 has an interpretation similar to that of Proposition 1 and implies that, under Assumption 4, any feasible solution of $\widetilde{\mathcal{P}}_3$ is accompanied by a probabilistic certificate regarding the satisfaction of the chance constraint in $\mathcal{P}_1$. Moreover, the number of samples that need to be extracted depends now on the number of constraints and not on the number of decision variables or the dimension of the uncertainty as in the standard scenario approach or Method 1, respectively. Note that Method 2 allows one to tackle problems where Method 1, though applicable, does not lead to a tractable robust problem. This is for example the case when $q_j(\delta)$ does not satisfy Assumptions 2 or 3. Moreover, there are cases where even if both Method 1 and 2 lead to a tractable problem, using Method 2 is of advantage since it leads to a less conservative performance.

If the scheme of Section III-B1 is employed, any problem (possibly non-convex) that exhibits the structure of Assumption 4 can be addressed by the proposed framework. On the other hand, if $\widetilde{\mathcal{P}}_3$ can be cast in the class of problems described in Section III-B2 then its robust counterpart can be transformed to a form that can be solved for instances of realistic size and its size is the same with the one reported in Section III-B2. An additional feature of the case addressed in this section is that in $\widetilde{\mathcal{P}}_3$ we treat each function $q_j(\delta)$ as an uncertainty input, therefore the constraint functions are linear with respect to $q_j(\delta)$ and Assumptions 2 and 3 are always satisfied.

V. Discussion and Numerical Results

A. Explicit Sample Complexity Bounds

Following [16], [18], [24], from (1) and for a given $\epsilon,\beta \in (0,1)$ explicit bounds for the sample size complexity can be obtained. Following Theorem 4 of [24], for the class of problems described in Section II which has $n = n_x$ decision variables, it suffices to generate

$$N \geq N_{\epsilon,\beta}^{n_x} = \left\lceil \frac{1}{\epsilon}\frac{e}{e-1}\left(n_x - 1 + \ln\frac{1}{\beta}\right)\right\rceil \tag{6}$$

samples to achieve the desired probabilistic performance. The operator $\lceil\cdot\rceil$ denotes the smallest integer greater than or equal to its argument and e is the Euler number. We will now provide similar bounds for the number of samples of the problems discussed in Section III; for the problems of Section IV the bounds are the same if n_δ is substituted with n_m.

We solve each problem in $\mathcal{P}_2$ using the scenario approach. This requires generating in total $\max_{i=1,\ldots,n_\delta} N_i$ samples, using N_i of them for each individual problem. The optimal solution τ_i^* of each problem would violate the corresponding constraint at most by ϵ_i. This provides additional design freedom and allows us to introduce different levels of violation for each uncertainty element. However, unless there is some physical intuition, there is no known systematic way to trade-off the constants ϵ_i and β_i. Following [25], an obvious choice is to select the same violation level $\epsilon_i = \epsilon/n_\delta$ and confidence $\beta_i = \beta/n_\delta$ for all $i = 1,\ldots,n_\delta$. In this case $N = N_i$ for all $i = 1,\ldots,n_\delta$ and, since we have $n = 2$ decision variables for each problem in $\mathcal{P}_2'$ we need to generate

$$N \geq N_{\epsilon,\beta}^{n_\delta} = \left\lceil \frac{n_\delta}{\epsilon}\frac{e}{e-1}\left(1 + \ln\frac{n_\delta}{\beta}\right)\right\rceil \tag{7}$$

uncertainty samples and use all of them in all n_δ problems of $\mathcal{P}_2'$. This bound is always lower compared to the case where an uneven distribution of ϵ_i and β_i is used.

For a given ϵ and β, an alternative approach is to compute simultaneously bounds for all elements of the uncertainty vector. In this case $\mathcal{P}_2'$ will no longer be a family of n_δ problems, but a single program whose constraints would be $\delta_i^{(k)} \in [\underline{\tau}_i, \overline{\tau}_i]$ for all $k = 1,\ldots,N$ and all $i = 1,\ldots,n_\delta$, and its objective function would be the sum of the interval lengths (i.e. $\sum_{i=1}^{n_\delta}(\overline{\tau}_i - \underline{\tau}_i)$). That way, we would have a problem with $n = 2n_\delta$ and hence

$$N \geq N_{\epsilon,\beta}^{n_\delta} = \left\lceil \frac{1}{\epsilon}\frac{e}{e-1}\left(2n_\delta - 1 + \ln\frac{1}{\beta}\right)\right\rceil. \tag{8}$$

Clearly, (8) leads to a lower bound on the number of samples relative to (7). By inspection of (6), (8), it should be noted that for all problem instances with $2n_\delta < n_x$, Method 1 requires fewer scenarios relative to the standard scenario approach. Although choosing the approach that requires the fewest scenarios prevents us from over-sampling, it does not necessarily lead to a computationally simpler or less conservative problem. This depends on the structure (i.e., the number and type of decision variables and constraints) of the resulting robust problem. For the standard scenario approach, the number of decision variables remains equal to n_x, whereas the number of constraints is $n_m N$. On the other hand, both the approaches proposed here result in a robust program with interval bounds on the uncertainty, whose size is determined following the discussion of Section III-B. The implications of the proposed methodology on the conservatism of the resulting solution are discussed by means of the numerical example of Section V-C.

The scenario approach provides a general purpose methodology to solve the chance constrained problems $\mathcal{P}_2$ (respectively $\widetilde{\mathcal{P}}_2$). However, alternative techniques could be employed as well. For example, in $\mathcal{P}_2$ we could formulate different problems to identify the minimum and maximum value (interpreted in a probabilistic sense) of the elements of δ. This would give rise to a family of $2n_\delta$ problems each of them having only one decision variable, thus falling in the worst-case performance framework of [8]. In this case if we select $\epsilon_i = \epsilon/2n_\delta$ and $\beta_i = \beta/2n_\delta$ for all $i = 1,\ldots,2n_\delta$, following [8], it suffices to generate $N \geq \lceil \ln(1/\beta_i)/\ln(1/1-\epsilon_i)\rceil \approx \lceil (2n_\delta/\epsilon)\ln(2n_\delta/\beta)\rceil$ uncertainty realizations, which is more conservative compared to (7) and (8). Moreover, if we seek to bound simultaneously all elements of δ, the procedure of [8] is no longer applicable, since the optimization problem would involve more than one decision variable.

B. Extension to the Sampling-and-Discarding Approach and Comparison With Other Methods

The scenario approach results were extended in [18], [19] to the so called sampling-and-discarding approach. Specifically, given N samples of the uncertainty, r of them are eliminated according to some rule and $\mathcal{P}_1'$ is formulated with the remaining $N-r$ samples. Under the assumption that almost surely the solution of the resulting problem violates the removed constraints (so that the solution is improved), the implications of Theorem 1 remain unchanged with the difference that given $\epsilon,\beta \in (0,1)$, N and r are selected according to

$$\binom{n+r-1}{r}\sum_{k=0}^{n+r-1}\binom{N}{k}\epsilon^k(1-\epsilon)^{N-k} \leq \beta \tag{9}$$

with $n = n_x$. Explicit bounds for the sample size are also provided in [18], [19]. Any algorithm could be employed for the discarding part; since optimal constraint discarding is of combinatorial complexity, [18] discusses the complexity of a greedy approach and an approach based on the Lagrange multipliers associated with the constraint functions.

In Method 1 (similarly for Method 2) we can incorporate these results when using the scenario approach in problems $\mathcal{P}_2'$ where we seek a hyper-rectangular set B^* which encloses the uncertainty with certain probability. Employing (9) with $n = 2n_\delta$ in Proposition 1 allows us to construct a B^* with smaller volume, thus reducing the conservatism of the solution of $\mathcal{P}_3$. There are multiple ways to select which r samples to discard; however, based on our simulation study the largest improvement in the cost is achieved when a greedy approach is adopted. We first solve the problem with N constraints and identify the ones that are active for $\mathcal{P}_2'$, or in other words the samples that lie on the facets of B^*. We then remove the one which results in the hyper-rectangle that leads to the highest reduction in the objective value of $\mathcal{P}_3$. Typically, this step requires solving $2n_\delta$ (assuming no multiple samples on the same facet of B^*) robust optimization problems. We then proceed in the same way until r samples are removed. In contrast to the scenario approach, the size of the robust problem at every step of this procedure does not depend on N or r.

Methods based on statistical learning theory [9] can also be used to provide similar results for a certain class of problems. In particular, for problems with finite VC dimension (see [9]) sample size bounds with complexity of $O((1/\epsilon^2)\ln(1/\epsilon^2)\ln(1/\beta))$ can be obtained, which clearly scale worse than those achieved by the scenario approach. Less conservative bounds $(O((1/\epsilon)\ln(1/\epsilon)\ln(1/\beta)))$ are derived in [26] to bound the so called probability of one-sided constrained failure. In principle these bounds can be applied to non-convex problems with finite VC dimension. However, they depend on an upper bound of the VC dimension, which is not necessarily easy to determine. The methods proposed here circumvent this difficulty at the cost of imposing assumptions on the dependency of the constraints functions on the uncertainty. A comparison of the learning-theoretic bounds with those of the scenario approach with optimal constraint removal is carried out in [18], where it is shown that for problems with linear constraints the latter tends to be exponentially better as the number of samples increases.

C. Numerical Example

Consider the problem

$$\min_{x\in\mathbb{R}^{n_x}, y\in\mathbb{R}} \|x\|_1 + |y| \text{ subject to :}$$

$$\mathbb{P}\left(\delta\in\Delta \mid \max_{j=1,\ldots,n_m}\left(\left(a_j^T+\delta^T B_j\right)x + c_j^T\delta + y\right)\leq 0\right)\geq 1-\epsilon \tag{10}$$

where $\delta \in \mathbb{R}^{n_\delta}$ is normally distributed with zero mean and identity covariance matrix. We consider problem instances with $n_x = n_m = 1,\ldots,19$, $n_\delta = 1,\ldots,5$. For all $j = 1,\ldots,n_m$ the vectors $a_j \in \mathbb{R}^{n_x}$, $c_j \in \mathbb{R}^{n_\delta}$ and the matrix $B_j \in \mathbb{R}^{n_\delta\times n_x}$ have all of their elements uniformly distributed in $[-1, 1]$. The additional decision variable $y \in \mathbb{R}$ is added to ensure Assumption 1 is satisfied, leading to $n_x + 1$ decision variables. For each case we use the standard scenario approach with $N = N_{\epsilon,\beta}^{n_x+1}$ given by (6) and Method 1 with $N = N_{\epsilon,\beta}^{n_\delta}$ given by (8).

We compute for each case the empirical probability of constraint violation, using 10,000 uncertainty realizations (not including the ones used for the optimization procedure), repeating the entire process for 100 different multi-sample extractions, keeping a_j, c_j, B_j, $j = 1,\ldots,n_m$ constant for all uncertainty realizations and multi-samples.

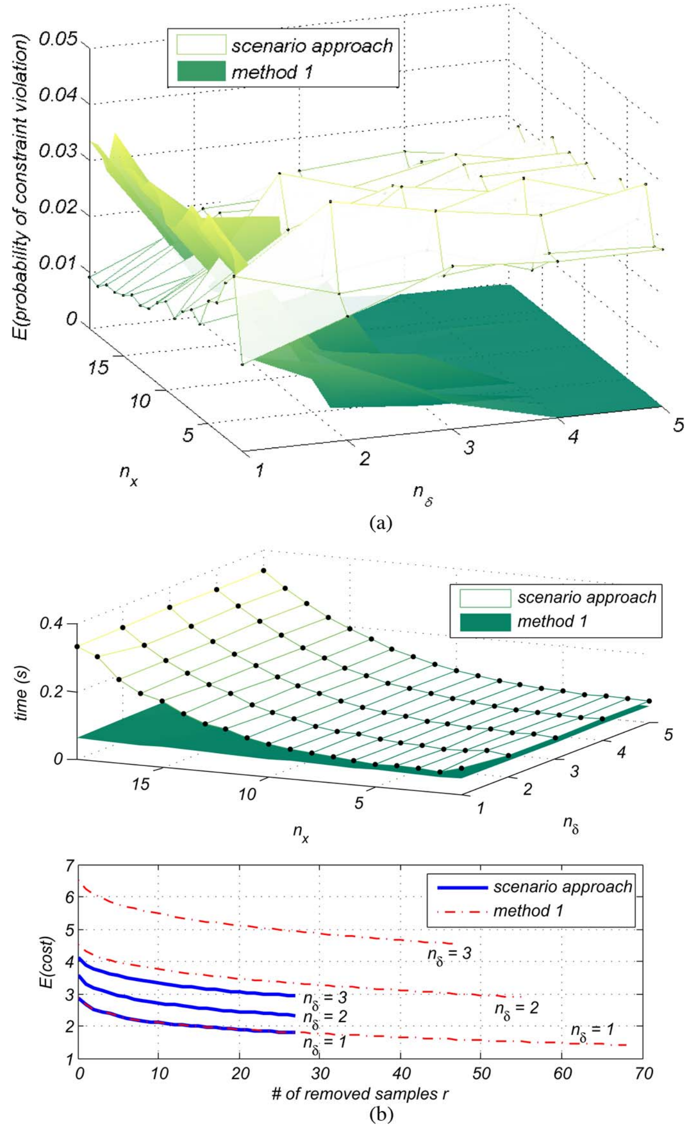

Fig. 1. (a) Expected empirical probability of constraint violation using Method 1 and the scenario approach with $\epsilon = 0.2$ and $\beta = 0.01$. (b) Upper panel: Average (over 100 multi-samples) computational time for the set-up of Fig. 1(a). Lower panel: Expected cost using the sampling-and-discarding approach both for Method 1 and the scenario approach with $\epsilon = 0.2$, $\beta = 0.01$ and $N = 500$ for instances of (10) with $n_x = 14$ and $n_\delta = 1, 2, 3$.

Fig. 1(a) shows the expected value of the empirical probability of constraint violation, which is always below the theoretical guarantees.

Consider first the case of scalar uncertainty, i.e. $n_\delta = 1$. Increasing the number of scenarios results in a more robust solution, which in turn leads to a lower violation probability. Hence, using the scenario approach, the probability of constraint violation decreases with respect to n_x, since the number of scenarios increases with the dimension of the decision vector. Since $N_{\epsilon,\beta}^{n_\delta}$ is independent of n_x, the probability of constraint violation does not change significantly with the number of decision variables. Moreover, for the case where $N_{\epsilon,\beta}^{n_\delta} < N_{\epsilon,\beta}^{n_x+1}$, our approach results in a higher violation probability, hence leads to a less conservative solution.

As n_δ increases, our approach becomes more conservative than the scenario approach. This is due to the fact that the solution of the scenario approach is guaranteed to be robust only with respect to $N_{\epsilon,\beta}^{n_x+1}$ uncertainty realizations, whereas with our approach the solution of $\mathcal{P}_3$ will also be robust with respect to all uncertainty realizations in B^*, not just the $N_{\epsilon,\beta}^{n_\delta}$ samples. This leads to low probabilities of constraint violation even in the case where $N_{\epsilon,\beta}^{n_\delta} < N_{\epsilon,\beta}^{n_x+1}$. Hence, for larger n_δ our approach is more conservative even though fewer uncertainty samples are required.

To compare our approach and the scenario approach in terms of cost, we compute for every problem instance the expected objective value (over the 100 multi-samples). The resulting cost surfaces follow a pattern similar to Fig. 1(a) with the roles of the scenario approach and Method 1 reversed, with the more conservative solution leading to the higher cost. The average cost difference increases with n_δ (for $n_\delta = 5$ the average cost increases by 100%).

Fig. 1(b) (upper panel) shows the average computational time for the set-up of Fig. 1(a). The scenario approach leads to higher average computational time compared to Method 1 for all problem instances. This difference tends to be much higher as n_x increases since the number of constraints in the scenario approach increases as well.

For the instances of (10) with $n_x = 14$ and $n_\delta = 1$, 2, 3 we compare Method 1 to the standard scenario approach for the case where the sampling-and-discarding approach of Section V-B is adopted in both methods. As shown in Fig. 1(b) (lower panel) the expected cost decreases monotonically (the expected probability of constraint violation increases with r). The evaluation is carried out against 10,000 uncertainty realizations and the expectation is with respect to 100 multi-samples. For a given ϵ, β and N, the number of removed constraints r is calculated from (9) using numerical inversion with $n = n_x + 1$ when using scenario approach and with $n = 2n_\delta$ when Method 1 is employed. For Method 1 we can remove more constraints since $2n_\delta < n_x + 1$ for the problem instances under consideration. Similarly to Fig. 1(a), for $n_\delta = 1$ Method 1 is less conservative than the scenario approach leading to lower cost. As n_δ increases, Method 1 leads to a more conservative performance compared to the scenario approach since a robust problem needs to be solved in our methodology. However, the size of this problem, and hence the computational time, is far lower compared to the one of the scenario approach.

All simulations were carried out using CPLEX [27] under the MATLAB interface YALMIP [28]. Application of our methodology to more realistic case studies can be found in [29].

VI. Conclusion

We proposed a methodology that eschews the direct application of the scenario approach to chance constrained optimization problems. It instead involves using the scenario approach in a lower dimensional, fully supported problem, constructing a subset of the uncertainty space. We then formulate the robust counterpart of the initial problem with the uncertainty confined in this set. Our approach provides guarantees with a reduced sample size, and does not require convexity as long as the resulting robust problem is solvable. We show that this is the case if one imposes assumptions on the way the constraint functions depend on the uncertainty, thus leading to a problem whose solution can be computed at a computational cost lower than the one of a scenario program. However, in cases where the uncertainty is of high dimension, the advantages of our solution come at the expense of a more conservative performance.

References

[1] A. Ben-Tal and A. Nemirovski, “Robust convex optimization,” *Math. Oper. Res.*, vol. 23, no. 4, pp. 769–805, 1998.

[2] L. E. Ghaoui and H. Lebret, “Robust solutions to uncertain semidefinite programs,” *SIAM J. Optim.*, vol. 9, no. 1, pp. 33–52, 1998.

[3] D. Bertsimas and M. Sim, “Tractable approximations to robust conic optimization problems,” *Math. Programming, Series B*, vol. 107, pp. 5–36, 2006.

[4] D. Bertsimas and M. Sim, “The price of robustness,” *Oper. Res.*, vol. 52, no. 1, pp. 35–53, 2004.

[5] A. Prekopa, *Stochastic Programming*. Boston, MA: Kluwer Academic Publishers, 1995.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[7] D. Bertsimas, D. Pachamanova, and M. Sim, “Robust linear optimization under general norms,” *Oper. Res. Lett.*, vol. 32, pp. 510–516, 2004.

[8] R. Tempo, E. W. Bai, and F. Dabbene, “Probabilistic robustness analysis: Explicit bounds for the minimum number of samples,” *Syst. Control Lett.*, vol. 30, no. 5, pp. 237–242, 1997.

[9] M. Vidyasagar, *A Theory of Learning and Generalization*. London, U.K.: Springer-Verlag, 1997.

[10] A. Lecchinni-Visintini, J. Lygeros, and J. Maciejowski, “Stochastic optimization on continuous domains with finite-time guarantees by Markov chain Monte Carlo methods,” *IEEE Trans. Autom. Control*, vol. 55, no. 12, pp. 2858–2863, 2010.

[11] A. Shapiro, “Stochastic programming approach to optimization under uncertainty,” *Math. Programming, Series B*, vol. 112, pp. 183–183, 2008.

[12] A. Nemirovski and A. Shapiro, “Probabilistic and Randomized Methods for Design under Uncertainty,” in *Scenario Approximation of Chance Constraints*, G. Calafiore and F. Dabbene, Eds. New York: Springer, 2005.

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

[14] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems*. London, U.K.: Springer-Verlag, 2005.

[15] G. Calafiore and M. Campi, “The scenario approach to robust control design,” *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, 2006.

[16] M. Campi and S. Garatti, “The exact feasibility of randomized solutions of uncertain convex programs,” *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.

[17] M. Campi and G. Calafiore, “Notes on the scenario design approach,” *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 382–385, 2009.

[18] G. Calafiore, “Random convex programs,” *SIAM J. Optim.*, vol. 20, no. 6, pp. 3427–3464, 1998.

[19] M. Campi and S. Garatti, “A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality,” *J. Optim. Theory Appl.*, vol. 148, no. 2, pp. 257–280, 2011.

[20] T. Alamo, R. Tempo, D. Ramirez, and E. Camacho, “A new vertex result for robustness problems with interval matrix uncertainty,” *Syst. Control Lett.*, vol. 57, no. 6, pp. 474–481, 2008.

[21] G. Calafiore and F. Dabbene, “Reduced vertex set result for interval semidefinite optimization problems,” *J. Optim. Theory Appl.*, vol. 139, no. 1, pp. 17–33, 2008.

[22] A. Ben-Tal and A. Nemirovski, “Robust solutions to uncertain programs,” *Oper. Res. Lett.*, vol. 25, no. 1, pp. 1–13, 1999.

[23] X. Lin, S. Janac, and C. Floudas, “A new robust optimization approach for scheduling under uncertainty: I. Bounded uncertainty,” *Comp. Chem. Eng.*, vol. 28, no. 6–7, pp. 1069–1085, 2004.

[24] T. Alamo, R. Tempo, and A. Luque, “On the sample complexity of randomized approaches to the analysis and design under uncertainty,” in *Proc. Amer. Control Conf.*, 2010, pp. 4671–4676.

[25] A. Nemirovski and A. Shapiro, “Convex approximations of chance constrained programs,” *Siam J. Control Optim.*, vol. 17, no. 4, pp. 969–996, 2006.

[26] T. Alamo, R. Tempo, and E. Camacho, “Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems,” *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2545–2559, 2009.

[27] *CPLEX11.0 Users Manual*, ILOG.SA., 2008, Technical report, Gentilly, France.

[28] J. Löfberg, “YALMIP: A toolbox for modeling and optimization in MATLAB,” in *Proc. IEEE Int. Symp. Comp. Aided Control Syst. Design*, 2005, pp. 284–289.

[29] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, “A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3885–3896, 2013.