Performance Bounds for the Scenario Approach and an Extension to a Class of Non-convex Programs

Peyman Mohajerin Esfahani, Tobias Sutter, and John Lygeros

Abstract—We consider the Scenario Convex Program (SCP) for two classes of optimization problems that are not tractable in general: Robust Convex Programs (RCPs) and Chance-Constrained Programs (CCPs). We establish a probabilistic bridge from the optimal value of SCP to the optimal values of RCP and CCP in which the uncertainty takes values in a general, possibly infinite dimensional, metric space. We then extend our results to a certain class of non-convex problems that includes, for example, binary decision variables. In the process, we also settle a measurability issue for a general class of scenario programs, which to date has been addressed by an assumption. Finally, we demonstrate the applicability of our results on a benchmark problem and a problem in fault detection and isolation.

1. Introduction

Optimization problems under uncertainty have considerable applications in disciplines ranging from mathematical finance to control engineering. For example most control systems involve some level of uncertainty; the aim of a robust control design is to provide a guaranteed level of performance for all admissible values of the uncertain parameters. In the convex case, two well-known approaches for dealing with such uncertain programs are robust convex programs (RCPs) and chance-constrained programs (CCPs). RCPs consider constraint satisfaction for all, possibly infinitely many, realizations of the uncertainty. While it is known that certain classes of RCPs can be solved as effectively as their non-robust counterparts [1] in other cases RCPs can be intractable [2], [3], [4], [5]. For example, the class of parametric linear matrix inequalities, which occur in many control problems, is NP-hard [6], [7]. CCPs, on the other hand, allow constraint violation with a low probability. The resulting optimization problem, however, is in general non-convex [8], [9].

Computationally tractable approximations to the aforesaid optimization problems can be obtained through the scenario convex programs (SCPs) in which only finitely many uncertainty samples are considered. A natural question in this case is how many samples would be "enough" to provide a good solution. To answer this question, one may view the problem from two perspectives: feasibility and objective performance. The literature mainly focuses on the first perspective. In this direction, the authors in [10], [11] initialized a feasibility

The authors are with the Automatic Control Laboratory, Physikstrasse 3, ETH Zürich, 8092 Zürich, Switzerland, {mohajerin,sutter,lygeros}@control.ee.ethz.ch

theory for CCP refined subsequently in [12], [13]. They established an explicit probabilistic lower bound for the sample size to guarantee the feasibility of the SCP solutions from a chance-constrained perspective. By contrast, the issue of performance bounds for both RCP and CCP via SCP has not been settled up to now. [14] provides a novel perspective in this direction that leads to optimal performance bounds for CCPs. However, it involves the problem of optimal constraint removal, which in general is computationally intractable.

The first contribution of this article is to address the SCP Performance issue from the objective viewpoint. The key element of our analysis relies on the concept of the worst-case violation inspired by the recent work [15]. The authors of [15] derived an upper bound of the worst-case violation for the SCPs where the uncertainty takes values in a finite dimensional Euclidean space. This result leads to a performance bound for a particular class of RCPs where the uncertainly appears in the objective function, e.g., min-max optimization problems. Motivated by different applications such as control problems with saturation constraints [16], fault detection and isolation in dynamical systems [17], and approximate dynamic programming [18], in this article we first extend this result to infinite dimensional uncertainty spaces. In the sequel, we establish a theoretical bridge from the optimal values of SCP to the optimal values of both RCP and CCP. Along this direction, under mild assumptions on the constraint function (measurability with respect to the uncertainty and lower semicontinuity with respect to the decision variables), we shall also rigorously settle a measurability issue of the SCP optimizer, which to date has been addressed in the literature by an assumption, e.g. [11], [12]. Our second contribution is to extend these results to a class of non-convex programs that, in particular, allows for binary decision variables. In the context of mixed integer programs, the recent work [19] investigates the feasibility perspective of CCPs, which leads to a bound of the required number of scenarios with exponential growth rate in the number of integer variables, whereas our proposed bound scales linearly.

The layout of this article is as follows: In Section 2 we formally introduce the optimization problems that will be addressed. Our results on probabilistic objective performance for both RCPs and CCPs based on SCPs are reported in Section 3. In Section 4 we extend our results to a class of non-convex programs, including mixed-integer programs with binary variables. To illustrate the proposed methodology, in Section 5 the theoretical results are applied to two examples: a benchmark problem whose solution can be computed explicitly, and a fault detection and isolation study with an

application to the security of power networks. We conclude in Section 6 with a summary of our work and comment on possible subjects of further research. For better readability, some of the technical proofs and details are given in the appendices.

NOTATION

Let $\mathbb{R}_+$ denote the non-negative real numbers. Given a metric space $\mathcal{D}$, its Borel σ-algebra is denoted by $\mathfrak{B}(\mathcal{D})$. Throughout this article, measurability always refers to Borel measurability. Given a probability space $\big(\mathcal{D}, \mathfrak{B}(\mathcal{D}), \mathbb{P}\big)$, we denote the N-Cartesian product set of $\mathcal{D}$ by $\mathcal{D}^N$ and the respective product measure by $\mathbb{P}^N$. An open ball in $\mathcal{D}$ with radius r and center v is denoted by $\mathbb{B}_r(v) := \{d \in \mathcal{D} : \|d - v\| < r\}$. The symbol $\models$ refers to the feasibility satisfaction, i.e., $x \models \mathrm{RCP}$ means that x is a feasible solution for the program RCP. Similarly, $x \not\models \mathrm{RCP}$ implies that x is not a feasible solution for the optimization problem RCP.

2. PROBLEM STATEMENT

Let $\mathbb{X} \subset \mathbb{R}^n$ be a compact convex set and $c \in \mathbb{R}^n$ a constant vector. Let $\big(\mathcal{D}, \mathfrak{B}(\mathcal{D}), \mathbb{P}\big)$ be a probability space where $\mathcal{D}$ is a metric space with the respective Borel σ-algebra $\mathfrak{B}(\mathcal{D})$. Consider the measurable function $f : \mathbb{X} \times \mathcal{D} \to \mathbb{R}$, which is convex in the first argument for each $d \in \mathcal{D}$, and bounded in the second argument for each $x \in \mathbb{X}$. We then consider the following optimization problems:

$$
\mathrm{RCP} : \begin{cases} \min\limits_{x} & c^\mathsf{T} x \\ \mathrm{s.t.} & f(x,d) \leq 0, \quad \forall d \in \mathcal{D} \\ & x \in \mathbb{X} \end{cases}
$$

$$
\mathrm{CCP}_\varepsilon : \begin{cases} \min\limits_{x} & c^\mathsf{T} x \\ \mathrm{s.t.} & \mathbb{P}[f(x,d) \leq 0] \geq 1 - \varepsilon \\ & x \in \mathbb{X} \end{cases} \tag{1}
$$

where $\varepsilon \in [0,1]$ is the constraint violation level for the chance-constrained program. We denote the optimal value of the program RCP (resp. CCP_ε) by $J^\star_{\mathrm{RCP}}$ (resp. $J^\star_{\mathrm{CCP}_\varepsilon}$). Suppose $(d_i)_{i=1}^N$ are N independent and identically distributed (i.i.d.) samples drawn according to the probability measure $\mathbb{P}$. The centerpiece of this study is the scenario program

$$
\mathrm{SCP} : \begin{cases} \min\limits_{x} & c^\mathsf{T} x \\ \mathrm{s.t.} & f(x, d_i) \leq 0, \quad \forall i \in \{1, \cdots, N\} \\ & x \in \mathbb{X} \end{cases} \tag{2}
$$

where the optimal solution and optimal value of SCP are denoted, respectively, by $x^\star_N$ and $J^\star_N$. Notice that SCP is naturally random as it depends on the random samples $(d_i)_{i=1}^N$.

We assume throughout our subsequent analysis that the following measurability assumption holds, though we shall show in Subsection 3-C how one may rigorously address this issue without any assumption for a large class of optimization programs (not necessarily convex).

Assumption 2.1. *The* SCP *optimizer generates a Borel measurable mapping from* $\big(\mathcal{D}^N, \mathfrak{B}(\mathcal{D}^N)\big)$ *to* $\big(\mathbb{X}, \mathfrak{B}(\mathbb{X})\big)$ *that associates each* $(d_i)_{i=1}^N$ *with a unique* $x^\star_N$.

The optimization program SCP in (2) is convex and hence tractable even for cases where the problems (1) are NP-hard. Motivated by this, a natural question is whether there exist theoretical links from SCP to RCP and CCP_ε. As mentioned in the introduction, this question can be addressed from two different perspectives: feasibility and objective performance. From the feasibility perspective, we recall the explicit bound of [12] which measures the finite sample behavior of SCP:

Theorem 2.2 (CCP_ε Feasibility). *Let* $\beta \in [0,1]$ *and* $N \geq N(\varepsilon, \beta)$ *where*

$$
N(\varepsilon, \beta) := \min \left\{ N \in \mathbb{N} \;\Big|\; \sum_{i=0}^{n-1} \binom{N}{i} \varepsilon^i (1 - \varepsilon)^{N-i} \leq \beta \right\}. \tag{3}
$$

Then, the optimizer of SCP *is a feasible solution of* CCP_ε *with probability at least* $1 - \beta$.

With "$\models$" notation, the assertion of Theorem 2.2 is alternatively stated by $\mathbb{P}^N\big[x^\star_N \models \mathrm{CCP}_\varepsilon\big] \geq 1 - \beta$, where $\mathbb{P}^N$ stands for the N-fold product probability measure.[1]

To the best of our knowledge, there is no clear connection between the feasibility of RCP and the solution of SCP. Furthermore, in Subsection 3-B we provide an example to challenge the possibility of such a connection. The focus of our study is on the second perspective to seek a (probabilistic) bound for the optimal values $J^\star_{\mathrm{RCP}}$ and $J^\star_{\mathrm{CCP}_\varepsilon}$ in terms of $J^\star_N$.

3. PROBABILISTIC OBJECTIVE PERFORMANCE

A. Confidence interval for the objective functions

The following definition inspired by the recent work [15] is the key object for our analysis.

Definition 3.1. *The* tail probability of the worst-case violation *is the function* $p : \mathbb{X} \times \mathbb{R}_+ \to [0,1]$ *defined as*

$$
p(x, \delta) := \mathbb{P}\Big[\sup_{v \in \mathcal{D}} f(x, v) - \delta < f(x, d) \Big].
$$

We call $h : [0,1] \to \mathbb{R}_+$ *a* uniform level-set bound *(ULB) of* p *if for all* $\varepsilon \in [0,1]$

$$
h(\varepsilon) \geq \sup \Big\{ \delta \in \mathbb{R}_+ \;\big|\; \inf_{x \in \mathbb{X}} p(x, \delta) \leq \varepsilon \Big\}.
$$

A pictorial representation of Definition 3.1 is given in Figure 1. Note that from a statistical perspective the ULB function may be alternatively viewed as an upper bound for the *quantile* function of the $\mathbb{R}$-valued random variable $d \mapsto \big(\sup_{v \in \mathcal{D}} f(x, v) - f(x, d)\big)$ uniformly in decision variable $x \in \mathbb{X}$ (cf. [20, Section 5.2]). Proposition 3.8 at the end of this subsection provides sufficient conditions under which a candidate ULB can be constructed. If the uncertainty set $\mathcal{D}$ is a specific compact subset of a Euclidean space, namely a norm-constrained or more generally a star-shaped set, the authors in [15] provide a constructive approach to obtain an admissible ULB.

---

[1]Note that $\mathbb{P}$ is the probability measure on $\mathfrak{B}(\mathcal{D})$; for simplicity we slightly abuse the notation, and will be doing so hereinafter. Strictly speaking, one has to define a new probability measure, say $\mathbb{Q}$, which is the induced measure on $\mathfrak{B}(\mathbb{X})$ via the mapping introduced in Assumption 2.1.
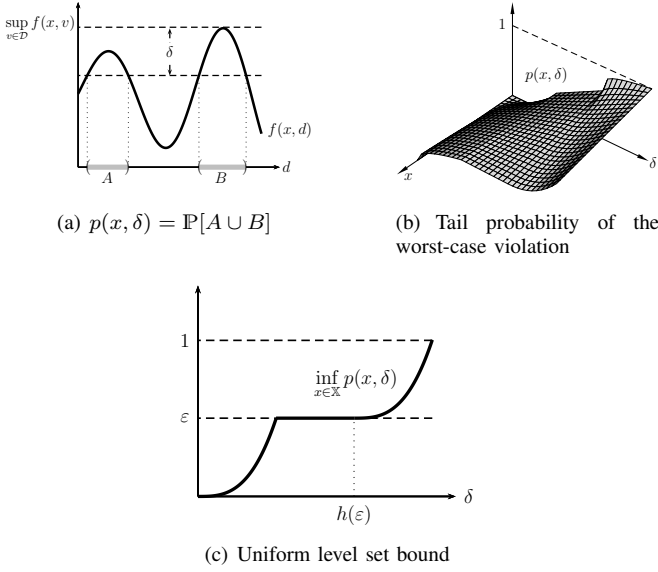
(a) $p(x, \delta) = \mathbb{P}[A \cup B]$

(b) Tail probability of the worst-case violation

(c) Uniform level set bound

Fig. 1.   Pictorial representation of Definition 3.1

Consider the relaxed version of the program RCP for $\gamma > 0$:

$$\mathrm{RCP}_\gamma : \begin{cases} \min_{x} & c^\mathsf{T} x \\ \text{s.t.} & f(x, d) \leq \gamma, \quad \forall d \in \mathcal{D} \\ & x \in \mathbb{X} \end{cases}, \qquad (4)$$

with the optimal value $J^\star_{\mathrm{RCP}_\gamma}$.

Lemma 3.2. *Let $h : [0, 1] \to \mathbb{R}_+$ be a ULB. Then,*

$$x \models \mathrm{CCP}_\varepsilon \quad \Longrightarrow \quad x \models \mathrm{RCP}_{h(\varepsilon)}$$

that is, the feasible set of the program CCP_ε with constraint violation level ε is a subset of the feasible set of the relaxed program RCP_γ with $\gamma := h(\varepsilon)$.

Proof: See Appendix A. ■

Assumption 3.3 (Slater Point). *There exists an $x_0 \in \mathbb{X}$ such that $\sup_{d \in \mathcal{D}} f(x_0, d) < 0$.*

Under Assumption 3.3, we define the constant

$$L_{\mathbf{SP}} := \frac{\min_{x \in \mathbb{X}} c^\mathsf{T} x - c^\mathsf{T} x_0}{\sup_{d \in \mathcal{D}} f(x_0, d)}. \qquad (5)$$

The following lemma is a classical result in perturbation theory of convex programs, which is a significant ingredient for the first result of this article.

Lemma 3.4. *Consider the relaxed program RCP_γ and its optimal value $J^\star_{\mathrm{RCP}_\gamma}$ as introduced in (4). Under Assumption 3.3, the mapping $\mathbb{R}_+ \ni \gamma \mapsto J^\star_{\mathrm{RCP}_\gamma} \in \mathbb{R}$ is Lipschitz continuous with constant bounded by $L_{\mathbf{SP}}$ in (5), i.e., for all $\gamma_2 \geq \gamma_1 \geq 0$ we have*

$$0 \leq J^\star_{\mathrm{RCP}_{\gamma_1}} - J^\star_{\mathrm{RCP}_{\gamma_2}} \leq L_{\mathbf{SP}}(\gamma_2 - \gamma_1).$$

Proof: See Appendix A. ■

Assumption 3.3 requires the existence of a strictly feasible solution x_0 which, in general, may not exist. However, in applications where a "risk-free" decision is available such

an assumption is not really restrictive; the portfolio selection problem is an example of this kind [21]. In addition, for the class of min-max problems, as a particular case of the program RCP, it is not difficult to see that Assumption 3.3 holds; see the following remark for more details and Section 5-B for an application to the problem of fault detection and isolation.

Remark 3.5 ($L_{\mathbf{SP}}$ for Min-Max Problems). *In min-max problems, one may inspect that there always exists a Slater point (in the sense of Assumption 3.3) with the corresponding constant $L_{\mathbf{SP}}$ arbitrarily close to 1. In fact, it is straightforward to observe that for min-max problems $J^\star_{\mathrm{RCP}_\gamma} = J^\star_{\mathrm{RCP}} - \gamma$, which readily implies that the Lipschitz constant of Lemma 3.4 is 1.*

The following results are the main contributions of the first part of the article.

Theorem 3.6 (RCP Confidence Interval). *Consider the programs RCP and SCP in (1) and (2) with the associated optimal values $J^\star_{\mathrm{RCP}}$ and $J^\star_N$, respectively. Suppose Assumption 3.3 holds and $L_{\mathbf{SP}}$ is the constant in (5). Given a ULB h and ε, β in $[0, 1]$, for all $N \geq N(\varepsilon, \beta)$ as defined in (3), we have*

$$\mathbb{P}^N\left[J^\star_{\mathrm{RCP}} - J^\star_N \in \left[0, I(\varepsilon)\right]\right] \geq 1 - \beta, \qquad (6)$$

where

$$I(\varepsilon) := \min\left\{L_{\mathbf{SP}} h(\varepsilon), \ \max_{x \in \mathbb{X}} c^\mathsf{T} x - \min_{x \in \mathbb{X}} c^\mathsf{T} x\right\}. \qquad (7)$$

Proof: Due to the definition of the optimization problems RCP and SCP, the second term of the confidence interval (7) is a trivial bound. It then suffices to establish the bound for the first term of (7). By Theorem 2.2, we know $\mathbb{P}^N\left[x^\star_N \models \mathrm{CCP}_\varepsilon\right] \geq 1 - \beta$ that in view of Lemma 3.2 implies

$$\mathbb{P}^N\left[x^\star_N \models \mathrm{RCP}_{h(\varepsilon)}\right] \geq 1 - \beta$$
$$\Longrightarrow \mathbb{P}^N\left[J^\star_{\mathrm{RCP}_{h(\varepsilon)}} \leq J^\star_N\right] \geq 1 - \beta,$$

where h is the ULB, and $J^\star_{\mathrm{RCP}_{h(\varepsilon)}}$ is the optimal value of the relaxed robust program (4) with $\gamma := h(\varepsilon)$. By virtue of Lemma 3.4, we have $J^\star_{\mathrm{RCP}} \leq J^\star_{\mathrm{RCP}_{h(\varepsilon)}} + L_{\mathbf{SP}} h(\varepsilon)$, that in conjunction with the above implication leads to

$$\mathbb{P}^N\left[J^\star_{\mathrm{RCP}} \leq J^\star_N + L_{\mathbf{SP}} h(\varepsilon)\right] \geq 1 - \beta.$$

Since the program SCP is just a restricted version of RCP, it is trivial that $J^\star_N \leq J^\star_{\mathrm{RCP}}$ pointwise on Ω^N, which concludes (6). ■

In accordance with the optimization problem CCP_ε, the following theorem provides similar performance assessment but in both a priori and a posteriori fashions.

Theorem 3.7 (CCP_ε Confidence Interval). *Consider the programs CCP_ε and SCP in (1) and (2) with the associated optimal values $J^\star_{\mathrm{CCP}_\varepsilon}$ and $J^\star_N$, respectively. Suppose Assumption 3.3 holds and $L_{\mathbf{SP}}$ is the constant in (5). Let h be a ULB and $\lambda^\star_N$ the dual optimizer of SCP. Given β in $[0, 1]$, for all $N \geq N(\varepsilon, \beta)$ defined in (3), we have*

A Priori Assessment:

$$\mathbb{P}^N\left[J^\star_{\mathrm{CCP}_\varepsilon} - J^\star_N \in \left[-I(\varepsilon), 0\right]\right] \geq 1 - \beta, \qquad (8a)$$

A Posteriori Assessment:

$$\mathbb{P}^N\Big[J^\star_{\mathrm{CCP}_\varepsilon} - J^\star_N \in \big[-I_N(\varepsilon), 0\big]\Big] \geq 1 - \beta, \qquad \text{(8b)}$$

where the a priori interval $I(\varepsilon)$ is defined as in (7), and the a posteriori interval is

$$I_N(\varepsilon) := \min\left\{\big\|\lambda^\star_N\big\|_1 h(\varepsilon),\ \max_{x\in\mathbb{X}} c^\mathsf{T}x - \min_{x\in\mathbb{X}} c^\mathsf{T}x\right\}. \qquad \text{(9)}$$

Proof: Similar to the proof of Theorem 3.6, we only need to show the first term of the confidence interval (9). In light of Theorem 2.2 and Lemma 3.2, we know that

$$\mathbb{P}^N\big[J^\star_{\mathrm{RCP}_{h(\varepsilon)}} \leq J^\star_{\mathrm{CCP}_\varepsilon} \leq J^\star_N\big] \geq 1 - \beta. \qquad \text{(10)}$$

In the same spirit as the previous proof, Lemma 3.4 ensures $J^\star_N \leq J^\star_{\mathrm{RCP}} \leq J^\star_{\mathrm{RCP}_{h(\varepsilon)}} + L_{\mathbf{SP}}h(\varepsilon)$ everywhere on Ω^N, which together with (10) arrives at (8a).

To show (8b), let us consider the scenario counterpart of the relaxed program RCP_γ in (4) with $\gamma := h(\varepsilon)$. We denote the optimal value of this scenario program by $J^\star_{N,h(\varepsilon)}$. Thus, we have $J^\star_{N,h(\varepsilon)} \leq J^\star_{\mathrm{RCP}_{h(\varepsilon)}}$ with probability 1. Notice that Assumption 3.3 also holds for the scenario program SCP, and consequently Lemma 3.4 is applicable to SCP as well. In fact, following the proof of Lemma 3.4 [22, p. 250], one can infer that the Lipschitz constant of the perturbation function can be over approximated by the ℓ_1-norm of a dual optimizer of the optimization program. Therefore, applying Lemma 3.4 to SCP yields to $J^\star_N - \|\lambda^\star_N\|_1 h(\varepsilon) \leq J^\star_{N,h(\varepsilon)} \leq J^\star_{\mathrm{RCP}_{h(\varepsilon)}}$ pointwise on Ω^N. Substituting into (10) leads to (8b). ∎

The parameter ε in Theorem 3.6 is a design choice which can be tuned to shrink the confidence interval $[0, I(\varepsilon)]$. On the contrary, in Theorem 3.7 the parameter ε is part of the problem data associated with the program CCP_ε. That is, in Theorem 3.7 $I(\varepsilon)$ is indeed fixed and the number of scenarios N in SCP only improves the confidence level β. In a same spirit but along a different approach, [13, Theorem 6.1] bounds $J^\star_N$ by the optimal solutions of two chance-constrained programs associated with different constraint violation levels, say $\bar\varepsilon < \varepsilon$. This value gap between the chance-constrained program and its scenario counterpart (either as explicitly derived in Theorem 3.7 or implicitly by two chance-constrained programs in [13, Theorem 6.1]) represents an inherent difference. To arbitrarily reduce the gap for CCP_ε, one may resort to *optimally* discarding a fraction of scenarios, which is in general computationally intractable; see for example [14, Theorem 6.1] and [13, Theorem 6.2].

By virtue of Theorem 3.6, the gap between $J^\star_{\mathrm{RCP}}$ and $J^\star_N$ is effectively quantified by a ULB $h(\varepsilon)$ as introduced in Definition 3.1. To control the behavior of $h(\varepsilon)$ as $\varepsilon \to 0$, one may require more structure on the measure $\mathbb{P}$ defined on $\big(\mathcal{D}, \mathfrak{B}(\mathcal{D})\big)$. Proposition 3.8 addresses this issue by introducing sufficient conditions concerning the measure of open balls in $\mathfrak{B}(\mathcal{D})$ and the continuity of the constraint mapping in the uncertainty argument.

Proposition 3.8. *Assume that the mapping $\mathcal{D} \ni d \mapsto f(x, d) \in \mathbb{R}$ is Lipschitz continuous with constant L_d uniformly in $x \in \mathbb{X}$. Suppose there exists a strictly increasing function $g : \mathbb{R}_+ \to [0, 1]$ such that*

$$\mathbb{P}\big[\mathbb{B}_r(d)\big] \geq g(r), \qquad \forall d \in \mathcal{D},$$

where $\mathbb{B}_r(d) \subset \mathcal{D}$ is an open ball centered at d with radius r. Then, $h(\varepsilon) := L_d\, g^{-1}(\varepsilon)$ is a ULB in the sense of Definition 3.1, where g^{-1} is the inverse function of g.

Proof: See Appendix A. ∎

Proposition 3.8 generalizes the corresponding result of [15, Lemma 3.1] by allowing the uncertainty space $\mathcal{D}$ to be possibly an infinite dimensional space. Note that the required assumptions in Proposition 3.8 implicitly require $\mathcal{D}$ to be bounded, though in practice this may not be really restrictive.

Remark 3.9. *Two remarks regarding the function g in Proposition 3.8 are in order:*

*(i) **Explicit expression:** Under the hypotheses of Proposition 3.8, Theorem 3.6 can be expressed in more explicit form. Let ε and β be in $[0, 1]$, L_d be the Lipschitz constant of the constraint function f in d, $L_{\mathbf{SP}}$ be the constant (5), and $N(\cdot, \cdot)$ be as defined in (3). Then, for any $N \geq N\big(g(\frac{\varepsilon}{L_{\mathbf{SP}}L_d}), \beta\big)$ we have*

$$\mathbb{P}^N\Big[J^\star_{\mathrm{RCP}} - J^\star_N \in [0, \varepsilon]\Big] \geq 1 - \beta.$$

*(ii) **Curse of dimensionality:** For an n_d-dimensional uncertainty set $\mathcal{D}$, the number of disjoint balls in $\mathcal{D}$ with radius r grows proportional to r^{-n_d} as r decreases. Thus, the assumptions of Proposition 3.8 imply that $g(r)$ is of the order of r^{n_d}. Therefore, for the desired precision ε, as detailed in the preceding remark, the required number of samples N grows exponentially as ε^{-n_d}. This appears to be an inherent feature when one seeks to bound the optimal value via scenario programs; see [23], [24] for similar observations.*

B. Feasibility of RCP via SCP

In this subsection we provide an example to show the inherent difficulty of the feasibility connection from SCP to the original problem RCP. Consider the following RCP with its SCP counterpart in which both decision and uncertainty space are compact subsets of $\mathbb{R}$:

$$\begin{cases} \min\limits_{x} & -x \\ \text{s.t.} & x - d \leq 0, \quad \forall d \in \mathcal{D} := [0, 1] \\ & x \in \mathbb{X} := [-1, 1] \end{cases},$$

$$\begin{cases} \min\limits_{x} & -x \\ \text{s.t.} & x - d_i \leq 0, \quad \forall i \in \{1, \cdots, N\} \\ & x \in \mathbb{X} := [-1, 1] \end{cases}.$$

It is not difficult to see that the feasible set of the robust program is $[-1, 0]$ with the optimizer $x^\star = 0$, whereas the optimizer of its scenario program is $x^\star_N = \min_{i \leq N} d_i$. If the probability measure $\mathbb{P}$ does not have atoms (point measure), we have $\mathbb{P}^N\big[\min_{i \leq N} d_i > 0\big] = 1$. Thus, one can deduce that

$$\mathbb{P}^N\big[x^\star_N \models \mathrm{RCP}\big] = 0, \qquad \forall \mathbb{P} \in \mathcal{P}, \quad \forall N \in \mathbb{N},$$

where $\mathcal{P}$ is the family of all nonatomic measures on $\big(\mathcal{D}, \mathfrak{B}(\mathcal{D})\big)$. More generally, if the set $\arg\max_{d\in\mathcal{D}} f(x, d)$ has measure zero for any $x \models \mathrm{RCP}$ (e.g., when f is convex in d and the boundary of $\mathcal{D}$ has zero measure), then the program SCP will almost surely return infeasible solutions

to the program RCP, as the worst-case scenarios are almost surely neglected.

C. Measurability of the SCP optimizer

The objective of this subsection is to address the standing Assumption 2.1. The measurability of the optimizer $x_N^\star$ for the scenario program SCP is a rather involved technical issue. In fact, to the best of our knowledge, in the literature this issue is always resolved by introducing an assumption. Let us highlight that the measurability of optimal values and the set of optimizers as well as the existence of a measurable selection are classical results in this context, see for instance [25, Theorem 14.37, p. 664]. However, there is no a priori guarantee that the obtained optimizer of the program SCP can be viewed as a measurable mapping from $\mathcal{D}^N$ to $\mathbb{X}$. Toward this issue, we propose a "two-stage" optimization program, in the *lexicographic* sense in the context of multi-objective optimization problems [26], in which the measurability of this mapping is ensured for a large class of programs (not necessarily convex).

For the rest of this section we assume that $\mathbb{X} \subset \mathbb{R}^n$ is closed and the mapping $x \mapsto f(x, d)$ is lower semicontinuous. Consider the scenario program SCP as defined in (2) with the corresponding optimal value $J_N^\star$; SCP is assumed to be feasible with probability one. Given the same uncertainty samples $(d_i)_{i=1}^N$ as in SCP, we introduce the second program

$$\begin{cases} \min_x & \phi(x) \\ \text{s.t.} & f(x, d_i) \le 0, \quad \forall i \in \{1, \cdots, N\} \\ & c^\mathsf{T} x \le J_N^\star \\ & x \in \mathbb{X} \end{cases}, \quad (11)$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ is a strictly convex function. Let us denote the optimizer of the above program by $\widetilde{x}_N^\star$. It is straightforward to observe that $\widetilde{x}_N^\star$ is indeed an optimizer of the program SCP.

Proposition 3.10 (Measurability of the Optimizer). *Consider the sequential two-stage programs* SCP *and* (11), *with the optimizer* $\widetilde{x}_N^\star$ *for the latter program. Then, the mapping* $\mathcal{D}^N \ni (d_i)_{i=1}^N \mapsto \widetilde{x}_N^\star \in \mathbb{X}$ *is a singleton and measurable.*

Proof: See Appendix A along with some preparatory lemmas. ∎

The above two-stage program may be viewed as a tie-break rule [10] or a regularization procedure [13, Section 2.1], which was proposed to resolve the uniqueness property of the SCP optimizer. Proposition 3.10 indeed asserts that the same trick ensures the measurability of the optimizer as well.

Remark 3.11 (Measurability of the Feasible Set). *The measurability of the feasibility event* $\widetilde{x}_N^\star \models \mathrm{CCP}_\varepsilon$ *(equivalently the measurability of the mapping* $x \mapsto \mathbb{P}[f(x, d) \le 0]$*) is a straightforward consequence of Proposition 3.10 and Fubini's Theorem [27, Thm. 18.3, p. 234].*

4. EXTENSION TO A CLASS OF NON-CONVEX PROGRAMS

This section extends the results developed in Section 3-A to a class of non-convex problems. Consider a family of programs introduced in (1) in which the program data are indexed by k, i.e., $(\mathbb{X}_k, f_k, \varepsilon_k)_{k=1}^m$. We assume that each tuple $(\mathbb{X}_k, f_k, \varepsilon_k)$ satisfies the required conditions in Section 2 (i.e., $\mathbb{X}_k$ is a compact convex set and the mapping $x \mapsto f_k(x, d)$ is convex for every $d \in \mathcal{D}$), and the corresponding programs are denoted by $\mathrm{RCP}^{(\mathrm{k})}$ and $\mathrm{CCP}_{\varepsilon_\mathrm{k}}^{(\mathrm{k})}$ as defined in (1). Consider the following (non-convex) optimization problems:

$$\begin{aligned} \mathrm{RP} : & \begin{cases} \min_x & c^\mathsf{T} x \\ \text{s.t.} & x \models \bigcup_{k=1}^m \mathrm{RCP}^{(\mathrm{k})} \end{cases}, \\ \mathrm{CP} : & \begin{cases} \min_x & c^\mathsf{T} x \\ \text{s.t.} & x \models \bigcup_{k=1}^m \mathrm{CCP}_{\varepsilon_\mathrm{k}}^{(\mathrm{k})} \end{cases}, \end{aligned} \quad (12)$$

where $x \models \bigcup_{k=1}^m \mathrm{RCP}^{(\mathrm{k})}$ $\big($resp. $x \models \bigcup_{k=1}^m \mathrm{CCP}_{\varepsilon_\mathrm{k}}^{(\mathrm{k})}\big)$ indicates that there exists $k \in \{1, \cdots, m\}$ such that $x \models \mathrm{RCP}^{(\mathrm{k})}$ $\big($resp. $x \models \mathrm{CCP}_{\varepsilon_\mathrm{k}}^{(\mathrm{k})}\big)$. In other words, the programs (12) seek an optimal solution which is feasible for at least one of the subprograms indexed by k, while the uncertainty space $\mathcal{D}$ as well as the associated measure $\mathbb{P}$ is shared between all the subprograms. Similarly, given i.i.d. samples $(d_i)_{i=1}^N \subset \mathcal{D}$ with respect to the probability measure $\mathbb{P}$, consider the scenario (non-convex) program

$$\mathrm{SP} : \begin{cases} \min_x & c^\mathsf{T} x \\ \text{s.t.} & x \models \bigcup_{k=1}^m \mathrm{SCP}^{(\mathrm{k})} \end{cases}. \quad (13)$$

Each subprogram $\mathrm{SCP}^{(\mathrm{k})}$ is defined according to the scenario convex program (2) associated with the program data $(\mathbb{X}_k, f_k)$ while the uncertainty samples $(d_i)_{i=1}^N$ are the same for all $k \in \{1, \cdots, m\}$. Before proceeding with the main result of this section, let us point out that the programs (12) contain, for example, a class of mixed integer programs. Let $f : \mathbb{R}^n \times \{0, 1\}^\ell \times \mathcal{D} \to \mathbb{R}$ be the constraint function in (1). It is straightforward to see that a chance-constrained mixed integer program can be formulated as

$$\begin{cases} \min_{x,y} & c^\mathsf{T} x \\ \text{s.t.} & \mathbb{P}[f(x, y, d) \le 0] \ge 1 - \varepsilon \\ & x \in \mathbb{X}, \quad y \in \{0, 1\}^\ell \end{cases}$$

$$\iff \begin{cases} \min_x & c^\mathsf{T} x \\ \text{s.t.} & \max_{k \in \{1, \cdots, 2^\ell\}} \mathbb{P}[f_k(x, d) \le 0] \ge 1 - \varepsilon \\ & x \in \mathbb{X} \end{cases},$$

where $f_k(x, d) := f(x, y_k, d)$ for each selection of the binary variables $y_k \in \{0, 1\}^\ell$. Then, by setting $m := 2^\ell$, $\mathbb{X}_k := \mathbb{X}$, $\varepsilon_k := \varepsilon$, the right-hand side of the above relation is readily in the framework of (12). A similar argument also holds for the robust mixed integer problems counterparts.

As a first step, we extend the feasibility result of Theorem 2.2 to the non-convex setting in (12).

Theorem 4.1 (CP Feasibility). *Let* $\vec{\varepsilon} := (\varepsilon_1, \cdots, \varepsilon_m)$, $\beta \in (0, 1]$, *and* $N \ge \widetilde{N}(\vec{\varepsilon}, \beta)$ *where*

$$\widetilde{N}(\vec{\varepsilon}, \beta) := \min\left\{ N \in \mathbb{N} \mid \sum_{k=1}^m \sum_{i=0}^{n-1} \binom{N}{i} \varepsilon_k^i (1 - \varepsilon_k)^{N-i} \le \beta \right\}. \quad (14)$$

Then, the optimizer of SP *is a feasible solution of* CP *with probability at least* $1 - \beta$.

Proof: Let $x_{N,k}^\star$ be the optimizer of the subprogram $\mathrm{SCP}^{(k)}$. By virtue of Theorem 2.2, one can infer that

$$\mathbb{P}^N\left[x_{N,k}^\star \not\models \mathrm{CCP}_{\varepsilon_k}^{(k)}\right] < \sum_{i=0}^{n-1}\binom{N}{i}\varepsilon_k^i(1-\varepsilon_k)^{N-i}.$$

On the other hand, it is straightforward to observe that the optimizer of the program SP, denoted by $x_N^\star$, belongs to the set $(x_{N,k}^\star)_{k=1}^m$. Therefore,

$$\mathbb{P}^N\left[x_N^\star \not\models \mathrm{CP}\right] \leq \mathbb{P}^N\left[\exists k \in \{1,\cdots,m\} \mid x_{N,k}^\star \not\models \mathrm{CCP}_{\varepsilon_k}^{(k)}\right]$$
$$\leq \sum_{k=1}^m \mathbb{P}^N\left[x_{N,k}^\star \not\models \mathrm{CCP}_{\varepsilon_k}^{(k)}\right]$$
$$< \sum_{k=1}^m\sum_{i=0}^{n-1}\binom{N}{i}\varepsilon_k^i(1-\varepsilon_k)^{N-i},$$

leading to the desired assertion. ∎

Remark 4.2 (Growth rate). *Notice that the number of subprograms,* m, *contributes to the confidence level* β *in a linear fashion. As an illustration, suppose* $\varepsilon_k := \varepsilon$. *In this case, one can easily verify that the confidence level of the non-convex program* SP *can be set equal to* $\frac{\beta}{m}$, *where* β *is the confidence level of each of the subprograms* $\mathrm{SCP}^{(k)}$. *From a computational perspective, one can follow the same calculation as in [28], and deduce that the contribution of* m *to the number of the required samples* $\widetilde{N}$ *appears in a logarithm. Thus, in our example of mixed integer programming above, the required number of samples grows linearly in the number of binary variables, which for most of applications could be considered a reasonable growth rate.*

The literature on computational schemes on non-convex problems is mainly based on statistical learning methods. A recent example of this nature is [29], which considers a class of problems involving Boolean expressions of polynomial functions. Given the degree and number of polynomial functions (α and k, respectively), the explicit sample bounds of [29] scale with $\varepsilon^{-1}\log(\alpha k\varepsilon^{-1})$ as opposed to our result in (14) which grows proportional to $\varepsilon^{-1}\log(m)$. We now proceed to extend the main results of Subsection 3-A, i.e., Theorems 3.6 and 3.7, to the non-convex settings (12) and (13) at once.

Theorem 4.3 (RP & CP Confidence Intervals). *Consider the programs* RP, CP, *and* SP *in* (12) *and* (13) *with the corresponding optimal values* $J_{\mathrm{RP}}^\star$, $J_{\mathrm{CP}}^\star$, *and* $J_N^\star$. *Given* $k \in \{1,\cdots,m\}$ *and the program data* $(\mathbb{X}_k, f_k)$, *let Assumption 3.3 hold and* $I^{(k)}$ *and* $I_N^{(k)}$ *be the a priori and a posteriori confidence intervals of the* k^{th} *subprogram as defined in* (7) *and* (9). *Then, given* $\beta \in [0,1]$ *and* $\vec{\varepsilon} := (\varepsilon_1, \cdots, \varepsilon_m) \in [0,1]^m$, *for all* $N \geq \widetilde{N}(\vec{\varepsilon}, \beta)$ *as defined in* (14) *we have*

A Priori Assessment:

$$\begin{cases} \mathbb{P}^N\left[J_{\mathrm{RP}}^\star - J_N^\star \in \left[0,\ \max_{k\leq m} I^{(k)}(\varepsilon)\right]\right] \geq 1-\beta, \\ \mathbb{P}^N\left[J_{\mathrm{CP}}^\star - J_N^\star \in \left[-\max_{k\leq m} I^{(k)}(\varepsilon),\ 0\right]\right] \geq 1-\beta, \end{cases}$$

A Posteriori Assessment:

$$\mathbb{P}^N\left[J_{\mathrm{CP}}^\star - J_N^\star \in \left[-\max_{k\leq m} I_N^{(k)}(\varepsilon),\ 0\right]\right] \geq 1-\beta.$$

Sketch of the proof: The proof effectively follows the same lines as in the proofs of Theorems 3.6 and 3.7. To adapt the required preliminaries, let us recall again that the optimizer of the programs (12) is one of the optimizers of the respective subprograms. The same assertion holds for the random program (13) as well. Moreover, since each subprogram of (13) fulfills the assumptions of Subsection 3-A, Lemmas 3.2 and 3.4 also hold for each subprogram with the corresponding data $(\mathbb{X}_k, f_k)$. Therefore, in light of Theorem 4.1, it only suffices to consider the worst-case possibility among all the subprograms. ∎

5. SIMULATION RESULTS

This section presents two examples to illustrate the theoretical results developed in the preceding sections and their performance. We first apply the results to a simple example whose analytical solution is available.

A. Example 1: Quadratic Constraint via Infinite Hyperplanes

Let $x = [x_1, x_2]^\mathsf{T}$ be the decision variables selected in the compact set $\mathbb{X} := [0,1]^2 \subset \mathbb{R}^2$, the linear objective function defined by $c := [-1, -1]^\mathsf{T}$, and the constraint function $f(x,d) := x_1\cos(d) + x_2\sin(d) - 1$ where the uncertainty d comes from the set $\mathcal{D} := [0, 2\pi]$. Consider the optimization problems introduced in (1) where $\mathbb{P}$ is the uniform probability measure on $\mathcal{D}$. It is not difficult to infer that the infinitely many hyperplane constraints can be replaced by a simple quadratic constraint. That is, for any $\gamma \geq 0$

$$\max_{d\in[0,2\pi]} x_1\cos(d) + x_2\sin(d) - 1 \leq \gamma$$
$$\iff \quad x_1^2 + x_2^2 \leq (\gamma+1)^2.$$

In the light of the above observation, we have the analytical solutions

$$\begin{aligned} J_{\mathrm{RCP}_\gamma}^\star &= \max\left\{-\sqrt{2}(\gamma+1), -2\right\}, \\ J_{\mathrm{CCP}_\varepsilon}^\star &= \max\left\{\frac{-\sqrt{2}}{\cos(\pi\varepsilon)}, -2\right\}, \end{aligned} \tag{15}$$

where $J_{\mathrm{RCP}_\gamma}^\star$ and $J_{\mathrm{CCP}_\varepsilon}^\star$ are the optimal values of the optimization problems RCP_γ and CCP_ε as defined in (4) and (1), respectively. The pictorial representation of the solutions is in Figure 2.

Let us fix the number of scenarios N for SCP in (2) with the optimal value $J_N^\star$. Given N and $\varepsilon \in [0,1]$, the confidence level $\beta \in [0,1]$ associated with our theoretical results is

$$\begin{aligned} \beta^*(\varepsilon) &:= \sum_{i=0}^{n-1}\binom{N}{i}\varepsilon^i(1-\varepsilon)^{N-i} \\ &= (1-\varepsilon)^N + N\varepsilon(1-\varepsilon)^{N-1}, \end{aligned}$$

where $n = 2$ in this example. Figure 3 depicts the behavior of $\beta^*(\varepsilon)$ for different values of N. Note that $x_0 = [0,0]^\mathsf{T}$ is a Slater point in the sense of Assumption 3.3 with the corresponding constant $L_{\mathrm{SP}} := \frac{-2-0}{-1} = 2$ (cf. (5)). Moreover,
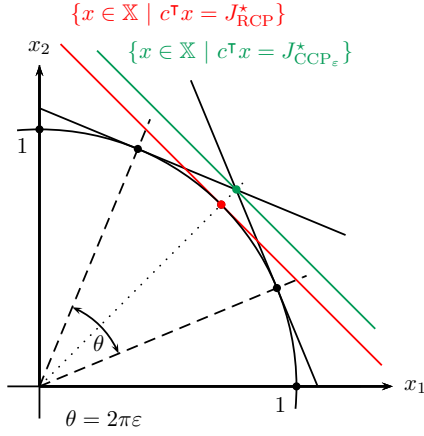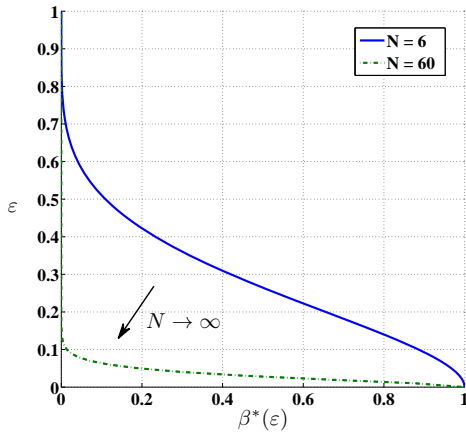
Fig. 2. Analytical solutions of Example 1

Fig. 3. Confidence level $\beta^*(\varepsilon)$ in terms of scenarios numbers N

it is easy to see that the mapping $d \mapsto f(x,d)$ has the Lipschitz constant $L_d = \sqrt{2}$ over the compact set $\mathbb{X} = [0,1]^2$. Thanks to Proposition 3.8 (and periodicity of the constraint function over the interval $[0, 2\pi]$), it is straightforward to introduce $g(r) = \frac{r}{\pi}$, and consequently obtain the ULB candidate $h(\varepsilon) := \sqrt{2}\pi\varepsilon$. Then, the confidence interval defined in (7) is $I(\varepsilon) := \max\{2\sqrt{2}\pi\varepsilon, 2\}$. As shown in Theorem 3.6 (resp. Theorem 3.7) we know that $J_{\mathrm{RCP}}^\star - J_N^\star \in [0, I(\varepsilon)]$ $\left(\text{resp. } J_{\mathrm{CCP}_\varepsilon}^\star - J_N^\star \in [-I(\varepsilon), 0]\right)$ with probability at least $1 - \beta^*(\varepsilon)$ for any $\varepsilon \in [0,1]$. To validate this result, we solve the program SCP for M different experiments. For each experiment $k \in \{1, \cdots, M\}$, we draw N scenarios $\left(d_i(k)\right)_{i=1}^N \subset [0, 2\pi]$ with respect to the uniform probability distribution $\mathbb{P}$ and solve the program SCP. Let $J_N^\star(k)$ be the optimal value of the kth experiment. Given $\beta \in [0,1]$, the empirical confidence interval of the program RCP can be represented by the minimal $\widetilde{I}(\beta)$ so that the interval $[0, \widetilde{I}(\beta)]$ contains $J_N^\star(m) - J_{\mathrm{RCP}}^\star$ for at least m experiments where $\frac{m}{M} \geq 1 - \beta$, i.e.,

$$\widetilde{I}(\beta) := \min\Big\{\widetilde{I} \in \mathbb{R}_+ \mid \exists A \subset \{1, \cdots, M\} :$$
$$|A| \geq (1 - \beta)M \quad \text{and}$$
$$J_{\mathrm{RCP}}^\star - J_N^\star(k) \in [0, \widetilde{I}] \ \forall k \in A\Big\}.$$

Regarding the program CCP_ε, notice that the empirical confidence interval depends on both parameters ε and β since the analytical optimal values $J_{\mathrm{CCP}_\varepsilon}^\star$ depends on ε as well. Hence, we define

$$\widetilde{I}_\varepsilon(\beta) := \min\Big\{\widetilde{I} \in \mathbb{R}_+ \mid \exists A \subset \{1, \cdots, M\} :$$
$$|A| \geq (1 - \beta)M \quad \text{and}$$
$$J_{\mathrm{CCP}_\varepsilon}^\star - J_N^\star(k) \in [-\widetilde{I}, 0] \ \forall k \in A\Big\}.$$
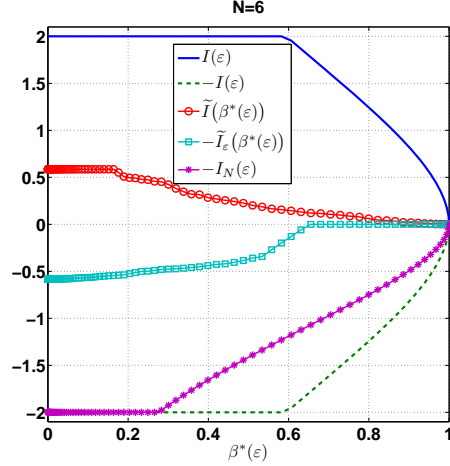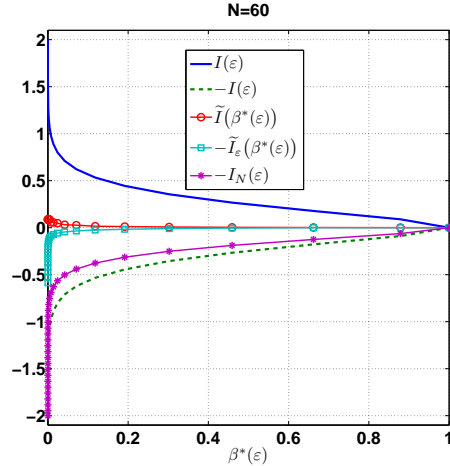

(a) Simulations for $N = 6$

(b) Simulations for $N = 60$

Fig. 4. Numerical results for Example 1

The sets $\widetilde{I}(\beta)$ and $\widetilde{I}_\varepsilon(\beta)$ are in close relation with sample quantiles in the sense of [20, Section 5.3.1]. In the following simulations the number of experiments is set to $M = 2000$. Figures 4(a) and 4(b) depict our theoretical performance bound $I(\varepsilon)$ for $N = 6$ and $N = 60$ in comparison with the empirical bounds $\widetilde{I}\left(\beta^*(\varepsilon)\right)$ and $\widetilde{I}_\varepsilon\left(\beta^*(\varepsilon)\right)$ where $\beta^*(\varepsilon)$ is the confidence level in Figure 3. As our theoretical results suggest, the confidence interval $[0, I(\varepsilon)]$ (resp. $[-I(\varepsilon), 0]$) contains the empirical interval $\left[0, \widetilde{I}\left(\beta^*(\varepsilon)\right)\right]$ (resp. $\left[-\widetilde{I}_\varepsilon\left(\beta^*(\varepsilon)\right), 0\right]$). Moreover, to demonstrate the a posteriori confidence interval in Theorem 3.7, we choose one of the experiments and depict the corresponding confidence interval $I_N(\varepsilon)$ versus $\beta^*(\varepsilon)$ as well. Note that in both cases of Figure 4 the a posteriori

confidence interval proposes a tighter bound than the a priori confidence interval. With this observation, we conjecture that in general the dual optimizer of SCP may happen to be a better approximation in comparison with the constant $L_{\mathbf{SP}}$ introduced in (5).

B. Example 2: Fault Detection and Isolation

The task of fault detection and isolation (FDI) involves generating a diagnosis signal to detect the occurrence of a specific fault. This is typically accomplished by designing a filter with all available signals as inputs (e.g., control signals and given measurements) and a scalar output that implements a non-zero mapping from the fault to the residual while decoupling unknown disturbances. In [30], a scalable optimization based approach is proposed to design an FDI filter for a class of nonlinear differential algebraic equation (DAE) where the filter is trained for finite number of disturbance signatures. The class of disturbances is further extended to a probability space in [17] where the filter performance is quantified in a probabilistic fashion.

As a particular subclass of DAEs, consider the nonlinear differential equation

$$\begin{cases} \dot{X}(t) = E\big(X(t)\big) + AX(t) + B_d d(t) + B_f f(t) \\ Y(t) = CX(t) \end{cases}, \quad (16)$$

where the matrices A, B_d, B_f, C and the function $E(\cdot)$ describe the linear and nonlinear dynamics of the model, respectively. Following [30], [17], we restrict the class of filters to linear transfer functions whose residual consists of two terms: $r = G[x](f) + r[x](d)$ where $G[x]$ is a linear time invariant transfer function expressing the mapping from the fault $f(\cdot)$ to the residual, and $r[x](d)$ is the contribution of the unknown disturbance $d(\cdot)$, and $x \in \mathbb{R}^n$ denotes the coefficients of the FDI filter to be designed. For linear systems (i.e., $E \equiv 0$) perfect decoupling between d and r may be possible (i.e., $r[x](d) \equiv 0$ for all d). For nonlinear systems, however, may not be the case. In this light, to minimize the impact of nonlinearities and disturbances on the residual, an optimal FDI filter can be obtained by the min-max program

$$\begin{cases} \min_{x,\gamma} & \gamma \\ \text{s.t.} & x^\mathsf{T} Q_d x \leq \gamma, \quad \forall d \in \mathcal{D} \\ & Hx = 0 \\ & \|Fx\|_\infty \geq 1 \end{cases}, \quad (17)$$

where the quadratic term $x^\mathsf{T} Q_d x$ represents the $\mathcal{L}_2$-norm of $r[x](d)$ over a given receding horizon, $\mathcal{D}$ is the space of possible disturbance patterns, and the last (non-convex) constraint is concerned with the norm of $G[x]$ as an operator. The matrices H and F are determined by the linear terms of the system dynamics (16), and the positive semidefinite matrix Q_d reflects the nonlinearity signature of the system dynamics in the presence of a disturbance pattern d; it depends on d and the nonlinear term $E(\cdot)$ of (16). We refer interested readers to [17] for details of the derivation of the above program.

For numerical case study, we consider an application of the above FDI design to detect a cyber intrusion in a two-area power network discussed in [30]. The setup in this example is a simplified version of [30, Section IV] where each power area contains one generator. Thus, the state in (16) comprises $X := \big[\Delta\phi, \{\Delta f_i\}_{1:2}, \{\Delta P_{m_i}\}_{1:2}, \{\Delta P_{agc_i}\}_{1:2}\big]^\mathsf{T}$ where $\Delta\phi$ is the voltage angle difference between the ends of the tie line, Δf_i the generator frequency, ΔP_{m_i} the generated mechanical power, and ΔP_{agc_i} the automatic generation control (AGC) signal in each area.[2]. The system dynamics is modeled in the framework of (16); the details are provided in Appendix B-A. The disturbance signal $d(\cdot)$ represents a load deviation that may occur in the first area. The signal f models the intrusion signal in the AGC of the first area, and the measurement signals are the frequencies and output power of the turbines, i.e., $Y = \big[\{\Delta f_i\}_{1:2}, \{\Delta P_{m_i}\}_{1:2}\big]^\mathsf{T}$. For a given horizon $T > 0$, we consider the class of disturbance signatures

$$\mathcal{D} := \Bigg\{ d : [0, T] \to \mathbb{R} \;\Big|\; \exists \alpha \in [0, 1],$$
$$d(t) := \sum_{k=0}^{p} a_k(\alpha) \cos\big(\frac{2\pi}{T} kt\big) \Bigg\},$$

where $a_k(\alpha)$ are the constant coefficients parametrized by α. The choice of $\mathcal{D}$ allows one to exploit available spectrum information of the disturbance signals. In this example, motivated by the emphasis on both low and high frequency regions, we assume $a_k(\alpha) := 5\big(\alpha 0.5^k + (1 - \alpha)0.5^{|10-k|}\big)$, $p = 30$, and $T = 4$ sec. For scenario generation, we consider a uniform probability distribution for the parameter $\alpha \in [0, 1]$, which in fact induces the probability measure $\mathbb{P}$ on $\mathcal{D}$. Let $d_0 \in \mathcal{D}$ be a disturbance signature with the corresponding parameter α_0. It is straightforward to observe that

$$\mathbb{P}\big[\|d - d_0\|_{\mathcal{L}_2} < r\big]$$
$$= \mathbb{P}\Big[\frac{T}{2} \sum_{k=0}^{p} \big|a_k(\alpha) - a_k(\alpha_0)\big|^2 < r^2\Big]$$
$$= \mathbb{P}\Bigg[|\alpha - \alpha_0| < \frac{\sqrt{2}\, r}{5\sqrt{T \sum_{k=0}^{p} \big(0.5^k - 0.5^{|10-k|}\big)^2}}\Bigg]$$
$$= \mathbb{P}\big[|\alpha - \alpha_0| < 0.142r\big] \geq 0.142r =: g(r),$$

where the function g, denoted in view of Proposition 3.8, is an invertible lower bound for the measure of open balls in $\mathcal{D}$. For the particular set of parameters in this example and specific operating region of interest, one can show that the mapping $d \mapsto Q_d$ is Lipschitz continuous with the constant $L_d = 0.02$; see Appendix B-B for more details. By virtue of Proposition 3.8 and normalizing[3] the optimizer of the SCP counterpart of the program (17), we can introduce the ULB candidate

$$h(\varepsilon) := L_d\, g^{-1}(\varepsilon) = 0.14\varepsilon.$$

Notice that the Infinite norm constraint in (17) is in fact a non-convex constraint. However, one may view it as the union of a finite number of constraint sets, see [30, Remark 3.2]. Therefore, the optimization problem (17) is already in the framework of RP as introduced in (12) where m is the number

---

[2]The symbol Δ stands for the deviation from the nominal value.
[3]Due to the linearity of the filter operator, one can always normalize the filter coefficients with no performance deterioration [17].

of rows in matrix F. It is remarkable that $m - 1$ equals the degree of the FDI filter chosen a priori. Thanks to the min-max structure of the robust program (17), the Lipschitz constant of Lemma 3.4 for each subprogram of (17) is $L_{\mathrm{SP}} = 1$, see Remark 3.5.

In this example, the dimension of the decision variable x is $n = 55$, the number of rows in F is $m = 5$, and the confidence level is set to $\beta = 0.01$. Therefore, to achieve the confidence interval $I(\varepsilon) = h(\varepsilon) = 5 \times 10^{-4}$, we need to set $\varepsilon = 3.57 \times 10^{-3}$ which, due to Theorem 4.1, requires to generate N disturbance signatures $d \in \mathcal{D}$ so that

$$N \geq \min\left\{ N \in \mathbb{N} \ \Big| \ \sum_{i=0}^{n-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \leq \frac{\beta}{m} \right\} = 22618.$$

Figures 5 demonstrate the numerical results of Example 2 over the course of 15 seconds. In Figure 5(a), 30 different realizations of disturbance inputs as well as an intrusion signal starting from $t = 10$ are shown in solid and dash curves, respectively. Figure 5(b) depicts the energy of the filter residual for the last $T = 4$ seconds (solid), and the threshold level associated with confidence $\beta = 0.01$ (dash). Notice that the proposed threshold is $\gamma^\star + 0.0005$, where $\gamma^\star$ is the optimal solution of the random counterpart of the program (17) with N = 22618 scenarios. Figure 5(c) presents the filter response which is the same figure as 5(b) but zoomed in on the period prior to the intrusion.
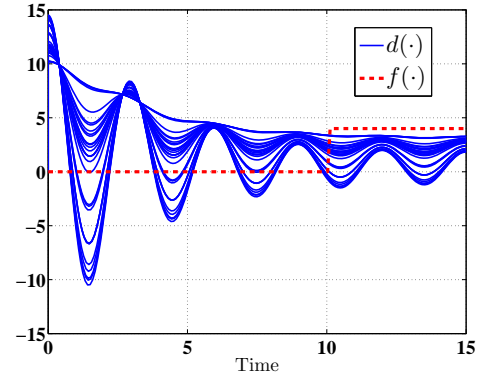
6. CONCLUSION AND FUTURE DIRECTION

In this article we presented probabilistic performance bounds for both RCP and CCP_ε via SCP. The proposed bounds are based on considering the tail probability of the worst-case constraint violation of the SCP solution as introduced in [15] together with some classical results from perturbation theory of convex optimization. In contrast to earlier approaches, this methodology is, to the best of our knowledge, the first confidence bounds for the objective performance of RCPs and CCPs based on scenario programs. Subsequently, we extended our results to a certain class of non-convex programs allowing for binary decision variables.

For future work, in light of Theorems 3.6 and 3.7, we aim to study the derivation of ULBs as introduced in Definition 3.1. Meaningful ULBs may depend highly on the individual structure of the optimization problems, in particular the uncertainty set and the constraint functions. Another potential direction, as highlighted by Example 1 in Section 5-A, is to investigate the relation between the constant L_{SP} in (5) and the dual optimizers of the program SCP.
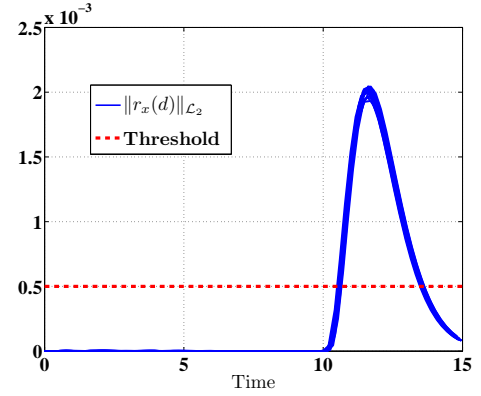
A. APPENDIX: TECHNICAL PROOFS

Proof of Lemma 3.2: Let h be a ULB as introduced in Definition 3.1, $x_0 \models \mathrm{CCP}_\varepsilon$, and $f^*(x_0) := \sup_{v \in \mathcal{D}} f(x_0, v)$. By definition of CCP_ε and p, the tail probability of the worst-case violation, we have
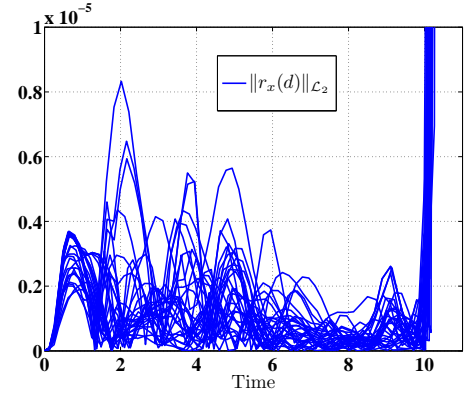
$$p\big(x_0, f^*(x_0)\big) \leq \varepsilon \implies \inf_{x \in \mathbb{X}} p\big(x, f^*(x_0)\big) \leq \varepsilon$$
$$\implies f^*(x_0) \leq h(\varepsilon) \implies x_0 \models \mathrm{RCP}_{h(\varepsilon)}$$

∎

(a) Scenarios of the disturbance signatures (solid), and intrusion signal (dash)

(b) Energy of the filter residual (solid), and the threshold level (dash)

(c) Residual response before the intrusion starts

Fig. 5. Numerical results for Example 2

Proof of Proposition 3.8: Given $x \in \mathbb{X}$, let $(v_i)_{i \in \mathbb{N}}$ be a sequence in $\mathcal{D}$ so that

$$\limsup_{i \in \mathbb{N}} f(x, v_i) = \sup_{v \in \mathcal{D}} f(x, v).$$

Thus, in light of Definition 3.1 we have

$$p(x, \delta) = \mathbb{P}\left[\sup_{v \in \mathcal{D}} f(x, v) - f(x, d) < \delta \right]$$
$$= \mathbb{P}\left[\limsup_{i \in \mathbb{N}} f(x, v_i) - f(x, d) < \delta \right]$$
$$\geq \mathbb{P}\left[\limsup_{i \in \mathbb{N}} L_d \|v_i - d\| < \delta \right] \quad \text{(A.1)}$$

$$\geq \limsup_{i\in\mathbb{N}} \mathbb{P}\left[\|v_i - d\| < \frac{\delta}{L_d}\right] \qquad (A.2)$$

$$= \limsup_{i\in\mathbb{N}} \mathbb{P}\left[\mathbb{B}_{\frac{\delta}{L_d}}(v_i)\right] \geq g\left(\frac{\delta}{L_d}\right),$$

where the first inequality in (A.1) follows from the Lipschitz continuity of f with respect to d, and the inequality in (A.2) is due to Fatou's lemma [31, p. 23]. Hence, in view of the ULB definition and the above analysis, we arrive at

$$\sup\left\{\delta\in\mathbb{R}_+ \mid \inf_{x\in\mathbb{X}} p(x,\delta)\leq\varepsilon\right\}$$

$$\leq \sup\left\{\delta\in\mathbb{R}_+ \mid g\left(\frac{\delta}{L_d}\right)\leq\varepsilon\right\} = L_d g^{-1}(\varepsilon).$$

∎

Proof of Lemma 3.4: It is well-known that under the strong duality condition the mapping $\gamma\mapsto \mathrm{RCP}_\gamma$, the so-called perturbation function, is Lipschitz continuous with the constant $\|\lambda^\star\|_1$ where $\lambda^\star$ is a dual optimizer of the program RCP; see [22, p. 250] for the proof and [32, Section 28] for more details in this direction. Now Lemma 3.4 follows from [33, Lemma 1], which essentially implies $\|\lambda^\star\|_1 \leq L_{\mathbf{SP}}$ where $L_{\mathbf{SP}}$ is the constant (5) corresponding to any Slater point in the sense of Assumption 3.3. ∎

To prove Proposition 3.10, we need some preliminaries.

Lemma A.1. *Let $\mathbb{X}\subset\mathbb{R}^n$ and $\mathcal{C}$ be the set of all lower semi-continuous functions from $\big(\mathbb{X},\mathfrak{B}(\mathbb{X})\big)$ to $\big(\mathbb{R},\mathfrak{B}(\mathbb{R})\big)$. Consider the mapping $J:\mathcal{C}\to\mathbb{R}$ defined by the optimization program*

$$\begin{cases} J(g) := & \min_x \quad c^\mathsf{T}x \\ & \mathrm{s.t.} \quad g(x)\leq 0 \\ & \qquad x\in\mathbb{X} \end{cases}. \qquad (A.3)$$

Then, the function J is measurable where the space of $\mathcal{C}$ is endowed with the infinite norm and the respective Borel σ-algebra.[4]

Proof: The proof is an application of [25, Theorem 14.37, p. 664]. Let us define the set-valued mapping $S:\mathcal{C}\rightrightarrows\mathbb{X}\times\mathbb{R}$ as follows:

$$S(g) := \big\{(x,\alpha)\in\mathbb{X}\times\mathbb{R} \mid \{g(x)\leq 0\}\ \&\ \{c^\mathsf{T}x\leq\alpha\}\big\}.$$

We first show that S is a normal integrand in the sense of [25, Definition 14.27, p. 661]. Since g is lower semicontinuous, then S is clearly closed-valued. We then only need to show that S is measurable according to [25, Definition 14.1, p. 643]. Let $O\subset\mathbb{X}\times\mathbb{R}$ be an open set, $(x_0,\alpha_0)\in O$ and $g_0\in S^{-1}(x_0,\alpha_0)$. Observe that for sufficiently small $\varepsilon>0$ we have $\mathbb{B}_\varepsilon(g_0)\subset S^{-1}(O)$ where $\mathbb{B}_\varepsilon(g_0) := \{g\in\mathcal{C} \mid \sup_{x\in\mathbb{X}}\|g(x)-g_0(x)\|\leq\varepsilon\}$, that implies that $S^{-1}(O)$ is open and in particular measurable. Thereby, S is measurable and hence a normal integral. Now the desired measurability readily follows from [25, Theorem 14.37, p. 664]. ∎

Lemma A.2. *Let $\phi:\mathbb{R}^n\to\mathbb{R}$ be a strictly convex function, and $\widetilde{J}:\mathcal{C}\to\mathbb{R}$ defined as follows:*

$$\begin{cases} \widetilde{J}(g) := & \min_x \quad \phi(x) \\ & \mathrm{s.t.} \quad g(x)\leq 0 \\ & \qquad c^\mathsf{T}x\leq J(g) \\ & \qquad x\in\mathbb{X} \end{cases}, \qquad (A.4)$$

where $J(g)$ is the function introduced in (A.3). Let $\widetilde{x}^\star(g)$ denote the set of optimizers of the program (A.4). Then, the mapping $\mathcal{C}\ni g\mapsto\widetilde{x}^\star\in\mathbb{R}^d$ is a measurable singleton.

Proof: Let us define the set-valued mapping $S:\mathcal{C}\rightrightarrows\mathbb{X}\times\mathbb{R}$

$$S(g) := \big\{(x,\alpha)\in\mathbb{X}\times\mathbb{R} \mid \{g(x)\leq 0\}$$
$$\&\ \{c^\mathsf{T}x - J(g)\leq 0\}\ \&\ \{\phi(x)\leq\alpha\}\big\}.$$

By virtue of the measurability of the mapping $g\mapsto J(g)$ in Lemma A.1 and along the same line of its proof, we know that S is a normal integral. Now, by [25, Theorem 14.37, p. 664] the existence of a measurable selection for the optimizer $\widetilde{x}^\star(g)$ as a function of $g\in\mathcal{C}$ is guaranteed. On the other hand, since $\phi:\mathbb{R}^n\to\mathbb{R}$ is strictly convex, the minimizer of the program (A.4) is unique. Therefore, $\widetilde{x}^\star(g)$ is a singleton and the desired measurability property follows at once. ∎

We now have all the required results to prove Proposition 3.10:

Proof of Proposition 3.10: Let $g:\mathcal{D}^N\to\mathcal{C}$ defined as

$$g(d_1,\cdots,d_N) := \max_{i\in\{1,\cdots,N\}} f(x,d_i). \qquad (A.5)$$

The measurability of the mapping (A.5) is ensured by the measurability assumption of the mapping $d\mapsto f(x,d)$ for each x. It is straightforward to observe that the optimizer of the program (11) can be viewed as the composition $\widetilde{x}_N^\star = \widetilde{x}^\star\circ g(d_1,\cdots,d_N)$ where $\widetilde{x}^\star$ is the optimizer of the program (A.4) and g is defined as in (A.5). Hence, the desired implication follows directly from the measurability of the mapping (A.5) and Lemma A.2. ∎

B. APPENDIX: DETAILS OF EXAMPLE 2

A. Mathematical model description

The two-area power network is described by the set of nonlinear ordinary differential equations

$$\Delta\dot{\phi} = 2\pi(\Delta f_1 - \Delta f_2),$$

$$\Delta\dot{f}_i = \frac{f_0}{2H_iS_{B_i}}\left(-\frac{1}{D_i}\Delta f_i - P_T\sin\Delta\phi + \Delta P_{m_i} - \Delta P_{load_i}\right),$$

$$\Delta\dot{P}_{m_i} = \frac{1}{T_{ch_i}}\left(-\frac{1}{S_i}\Delta f_i - \Delta P_{m_i} + \Delta P_{agc_i}\right),$$

$$\Delta\dot{P}_{agc_i} = \left(\frac{1}{D_i}\frac{C_if_0}{2S_iH_iS_{B_i}} - \frac{1}{S_i}\frac{1}{T_{N_i}}\right)\Delta f_i$$
$$\quad - \frac{C_if_0}{2S_iH_iS_{B_i}}(\Delta P_{m_i} - \Delta P_{load_i}) - \frac{C_if_0}{2S_iH_iS_{B_i}}\Delta P_{agc_i}$$
$$\quad - \left(\frac{1}{T_{N_i}} - \frac{C_if_0}{2S_iH_iS_{B_i}}\right)P_T\sin\Delta\phi$$
$$\quad - 2\pi C_i P_T(\Delta f_1 - \Delta f_2)\cos\Delta\phi,$$

where $i\in\{1,2\}$ is the index of each area, $X := [\Delta\phi, \{\Delta f_i\}_{1:2}, \{\Delta P_{m_i}\}_{1:2}, \{\Delta P_{agc_i}\}_{1:2}]^\mathsf{T}\in\mathbb{R}^7$ is the state vector, and the constant parameters in this example are chosen the

same for both areas as $T_{ch_i} = 5$ sec, $S_{B_i} = 1.8$ GW, $f_0 = 50$ Hz, $H_i = 6.5$ sec, $D_i = 428.6$ Hz/GW, $S_i = 1.389$ Hz/GW, $C_i = 0.1$, $T_{N_i} = 30$, $P_T = 0.15$ GW. We refer to [35] for physical interpretation of these parameters and more details on the model equations. In the example, we assume that $\Delta P_{load_1} = d$ where $d \in \mathcal{D}$ is the disturbance signal and $\Delta P_{load_2} \equiv 0$.

B. Lipschitz constant of the mapping $d \mapsto Q_d$

This mapping can be viewed in two steps: $d \mapsto E(X)$ and $E(X) \mapsto Q_d$ where X is the solution process in the presence of the disturbance input d, and E the nonlinear term of the ODE (16). The key step is to approximate the Lipschitz constant of the first mapping $d \mapsto E(X)$. The classical result of the continuity of the ODEs solution, obtained by Lipschitz continuity of the vector field and Gronwall's inequality, turns out to be too conservative in this case. We then invoke a Lyapunov-like approach to address this issue more efficiently. Let us define the shorthand $h(X, d) := E(X) + AX + B_d d$. Suppose there exist a function $V : \mathbb{R}^7 \times \mathbb{R}^7 \to \mathbb{R}_+$ and positive constants κ, ρ so that for every $X, \widetilde{X} \in \mathbb{R}^7$ and $d, \widetilde{d} \in \mathbb{R}$

$$\left\| E(X) - E(\widetilde{X}) \right\|^2 \leq V(X, \widetilde{X}) \tag{B.1a}$$

$$\partial_X V(X, \widetilde{X}) h(X, d) + \partial_{\widetilde{X}} V(X, \widetilde{X}) h(\widetilde{X}, \widetilde{d})$$
$$\leq -\kappa V(X, \widetilde{X}) + \rho |d - \widetilde{d}|. \tag{B.1b}$$

Using standard Gronwall's inequality, one can show that under conditions (B.1) we have

$$\left\| E(X) - E(\widetilde{X}) \right\|_{\mathcal{L}_2}^2 \leq \int_0^T \left\| E(X(t)) - E(\widetilde{X}(t)) \right\|^2 \mathrm{d}t$$
$$\leq \int_0^T V(X(t), \widetilde{X}(t)) \mathrm{d}t$$
$$\leq \rho \int_0^T \mathrm{e}^{-\kappa t} \int_0^t |d(s) - \widetilde{d}(s)| \mathrm{d}s \mathrm{d}t$$
$$\leq \frac{2\rho}{3} T \sqrt{T} \|d - \widetilde{d}\|_{\mathcal{L}_2}.$$

In [36, Theorem 3.3], a similar technique is discussed in more detail to establish a connection between the Lyapunov function and continuity of the solution trajectories. In order to find a Lyapunov function in the above sense, we limit our search domain to the quadratic functions, i.e., $V(X, \widetilde{X}) = (X - \widetilde{X})^\mathsf{T} Q (X - \widetilde{X})$ for some positive semidefinite matrix Q. It is not difficult to deduce that the nonlinear term E effectively depends only on the state $\Delta\phi$. Hence, to fulfill the requirement (B.1a) it suffices to guarantee $Q \succeq vv^\mathsf{T}$ where $v = [0, 0, 1, 0, 0, 0, 0]^\mathsf{T}$. Setting $\kappa = 0.01$, we then solve the set of linear matrix inequalities (LMIs)

$$\begin{cases} \min_{\sigma, Q} & \sigma \\ \text{s.t.} & QA^\mathsf{T} + AQ \preceq -\kappa Q \\ & vv^\mathsf{T} \preceq Q \preceq \sigma I \end{cases},$$

which provides a local Lyapunov function in the sense of (B.1). Note that one can always extract the linear part of E and

add it to the matrix A. Now, by numerical inspection, it turns out that for the specific system parameters of this example, V obtained from the above LMIs is a Lyapunov function in the domain of $\Delta f_i \in [-0.1, 0.1]$ Hz, $\Delta\phi \in [-10°, 10°]$, $\Delta p_{m_i} \in [-10, 10]$ MW, $\Delta p_{agc_i} \in [-15, 15]$ MW. Therefore, the parameter ρ in (B.1b) can be numerically approximated via the optimal σ in the LMIs together with matrix B_d and the region of interest described above. Besides, since the FDI filter is a stable linear time invariant transfer function with normalized coefficients, the Lipschitz constant of the second mapping $E(X) \mapsto Q_d$ can be explicitly computed based on the filter denominator which is fixed prior to the design procedure; see [17, Lemma 4.5].

REFERENCES

[1] D. Bertsimas and M. Sim, "Tractable approximations to robust conic optimization problems," *Mathematical Programming*, vol. 107, no. 1-2, pp. 5–36, 2006.

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

[3] ——, "Robust solutions of uncertain linear programs," *Operations Research Letters*, vol. 25, pp. 1–13, 1999.

[4] L. E. Ghaoui, F. Oustry, and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 33–52, 1998.

[5] A. Ben-tal, A. Nemirovski, and C. Roos, "Robust solutions of uncertain quadratic and conic-quadratic problems," in *Solutions of Uncertain Linear Programs: Math. Program.* Kluwer, 2001, pp. 351–376.

[6] S. Boyd, L. E. Ghaoul, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Society for Industrial and Applied Mathematics: SIAM studies in applied mathematics. Society for Industrial and Applied Mathematics, 1994.

[7] P. Gahinet, "Explicit controller formulas for lmi-based h-infinity synthesis," *Automatica*, vol. 32, no. 7, pp. 1007–1014, Jul. 1996.

[8] A. Prékopa, *Stochastic Programming*, ser. Mathematics and Its Applications. Springer, 1995.

[9] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, ser. MPS-SIAM series on optimization. Society for Industrial and Applied Mathematics, 2009.

[10] G. Calafiore and M. C. Campi, "Uncertain convex programs: randomized solutions and confidence levels," *Math. Program.*, vol. 102, no. 1, Ser. A, pp. 25–46, 2005. [Online]. Available: http://dx.doi.org/10.1007/s10107-003-0499-y

[11] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Automat. Control*, vol. 51, no. 5, pp. 742–753, 2006. [Online]. Available: http://dx.doi.org/10.1109/TAC.2006.875041

[12] M. C. 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. [Online]. Available: http://dx.doi.org/10.1137/07069821X

[13] G. C. Calafiore, "Random convex programs," *SIAM J. Optim.*, vol. 20, no. 6, pp. 3427–3464, 2010. [Online]. Available: http://dx.doi.org/10.1137/090773490

[14] M. C. 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. [Online]. Available: http://dx.doi.org/10.1007/s10957-010-9754-6

[15] T. Kanamori and A. Takeda, "Worst-case violation of sampled convex programs for optimization with uncertainty," *Journal of Optimization Theory and Applications*, vol. 152, no. 1, pp. 171–197, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10957-011-9923-2

[16] M. C. Campi, S. Garatti, and M. Prandini, "The scenario approach for systems and control design," *Annual Reviews in Control*, no. 2, pp. 149–157, 2009.

[17] P. Mohajerin Esfahani and J. Lygeros, "A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance," Tech. Rep., Feb. 2013, [Online]. Available: http://control.ee.ethz.ch/index.cgi?page=publications&action=details&id=4344.

[18] F. Dufour and T. Prieto-Rumeau, "Finite linear programming approximations of constrained discounted Markov decision processes," *SIAM J. Control Optim.*, vol. 51, no. 2, pp. 1298–1324, 2013. [Online]. Available: http://dx.doi.org/10.1137/120867925

[19] G. C. Calafiore, D. Lyons, and L. Fagiano, "On mixed-integer random convex programs," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 2012, pp. 3508–3513.

[20] J. Shao, *Mathematical statistics*, 2nd ed., ser. Springer Texts in Statistics. New York: Springer-Verlag, 2003. [Online]. Available: http://dx.doi.org/10.1007/b97553

[21] 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," *J. Optim. Theory Appl.*, vol. 155, no. 2, pp. 707–722, 2012. [Online]. Available: http://dx.doi.org/10.1007/s10957-012-0074-x

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[23] A. Lecchini-Visintini, J. Lygeros, and J. M. Maciejowski, "Approximate domain optimization for deterministic and expected value criteria," Tech. Rep., Mar. 2008, [Online]. Available: http://control.ee.ethz.ch/index.cgi?page=publications;action=details;id=3048.

[24] ——, "Stochastic optimization on continuous domains with finite-time guarantees by Markov chain Monte Carlo methods," *IEEE Trans. Automat. Control*, vol. 55, no. 12, pp. 2858–2863, 2010. [Online]. Available: http://dx.doi.org/10.1109/TAC.2010.2078170

[25] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, ser. Grundlehren der mathematischen Wissenschaften. Springer, 2010.

[26] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and multidisciplinary optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[27] P. Billingsley, *Probability and measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc., 1995, a Wiley-Interscience Publication.

[28] G. C. Calafiore, "A note on the expected probability of constraint violation in sampled convex programs," in *18th IEEE International Conference on Control Applications Part of 2009 IEEE Multi-conference on Systems and Control*, july 2009, pp. 1788 – 1791.

[29] T. Alamo, R. Tempo, and E. F. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," *IEEE Trans. Automat. Control*, vol. 54, no. 11, pp. 2545–2559, 2009. [Online]. Available: http://dx.doi.org/10.1109/TAC.2009.2031207

[30] P. Mohajerin Esfahani, M. Vrakopoulou, G. Andersson, and J. Lygeros, "A tractable nonlinear fault detection and isolation technique with application to the cyber-physical security of power systems," in *51th IEEE Conference Decision and Control*, 2012, [Online]. Full version: http://control.ee.ethz.ch/index.cgi?page=publications;action=details;id=4196.

[31] W. Rudin, *Real and complex analysis*, ser. Mathematics series. McGraw-Hill, 1987.

[32] R. T. Rockafellar, *Convex analysis*, ser. Princeton Landmarks in Mathematics and Physics Series. PRINCETON University Press, 1997.

[33] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1757–1780, 2008. [Online]. Available: http://dx.doi.org/10.1137/070708111

[34] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Nonlinear parametric optimization*. Basel: Birkhäuser Verlag, 1983.

[35] P. Mohajerin Esfahani, M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "Cyber attack in a two-area power system: Impact identification using reachability," in *American Control Conference*, 2010, pp. 962 – 967.

[36] M. Zamani, P. Mohajerin Esfahani, A. Abate, and J. Lygeros, "Symbolic models for stochastic control systems," *to appear in the IEEE Transactions on Automatic Control*, 2014, [Online]. Available: http://arxiv.org/abs/1302.3868.

Peyman Mohajerin Esfahani received the B.Sc. and M.Sc degrees in Electrical Engineering both from Sharif University of Technology, Tehran, Iran, in September 2005 and December 2007, respectively. In June 2008 he joined the Automatic Control Laboratory at the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, where he obtained the Ph.D. degree in January 2014.

He currently holds a dual-appointment as a postdoctoral researcher at the Risk Analytics and Optimization Chair at ETH Lausanne (EPFL) and the Automatic Control Laboratory at ETH Zurich. His research interests include control synthesis, analysis, and decision making problems in uncertain, adversarial, and dynamic environments subject to information constraints, with applications to large scale systems such as power networks and industrial cyber-physical systems.

Tobias Sutter received a B.Sc. degree in Mechanical Engineering in 2010 and an M.Sc. degree in Robotics, Systems and Control in 2012, both from ETH Zurich, Switzerland and is currently pursuing a Ph.D. degree in Electrical Engineering at the Automatic Control Laboratory at ETH Zurich.

His research interests include approximate dynamic programming, control under communication constraints, and information theory.

John Lygeros holds the chair of Computation and Control at the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, where he is currently serving as the Head of the Automatic Control Laboratory. He completed a B.Eng. degree in electrical engineering in 1990 and an M.Sc. degree in Systems Control in 1991, both at Imperial College of Science Technology and Medicine, London, U.K.. In 1996 he obtained a Ph.D. degree from the Electrical Engineering and Computer Sciences Department, University of California, Berkeley. During the period 1996-2000 he held a series of research appointments at the National Automated Highway Systems Consortium, Berkeley, the Laboratory for Computer Science, M.I.T., and the Electrical Engineering and Computer Sciences Department at U.C. Berkeley. Between 2000 and 2003 he was a University Lecturer at the Department of Engineering, University of Cambridge, U.K., and a Fellow of Churchill College. Between 2003 and 2006 he was an Assistant Professor at the Department of Electrical and Computer Engineering, University of Patras, Greece. In July 2006 he joined the Automatic Control Laboratory at ETH Zurich, first as an Associate Professor, and since January 2010 as Full Professor. His research interests include modeling, analysis, and control of hierarchical, hybrid, and stochastic systems, with applications to biochemical networks, automated highway systems, air traffic management, power grids and camera networks. John Lygeros is a Fellow of the IEEE, and a member of the IET and of the Technical Chamber of Greece.