Automatica 52 (2015) 160–172

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms☆

Teodoro Alamo [a], Roberto Tempo [b], Amalia Luque [a], Daniel R. Ramirez [a]

[a] Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla, Escuela Superior de Ingenieros, Camino de los Descubrimientos s/n, 41092 Sevilla, Spain

[b] CNR-IEIIT, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

ARTICLE INFO

Article history:
Received 18 March 2013
Received in revised form
18 July 2014
Accepted 27 October 2014
Available online 10 December 2014

Keywords:
Randomized and probabilistic algorithms
Uncertain systems
Sample complexity

ABSTRACT

In this paper, we study randomized methods for feedback design of uncertain systems. The first contribution is to derive the sample complexity of various constrained control problems. In particular, we show the key role played by the binomial distribution and related tail inequalities, and compute the sample complexity. This contribution significantly improves the existing results by reducing the number of required samples in the randomized algorithm. These results are then applied to the analysis of worst-case performance and design with robust optimization. The second contribution of the paper is to introduce a general class of sequential algorithms, denoted as Sequential Probabilistic Validation (SPV). In these sequential algorithms, at each iteration, a candidate solution is probabilistically validated, and corrected if necessary, to meet the required specifications. The results we derive provide the sample complexity which guarantees that the solutions obtained with SPV algorithms meet some pre-specified probabilistic accuracy and confidence. The performance of these algorithms is illustrated and compared with other existing methods using a numerical example dealing with robust system identification.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The use of randomized algorithms for systems and control has matured thanks to the considerable research efforts made in recent years. Key areas where we have seen convincing developments include uncertain and hybrid systems (Tempo, Calafiore, & Dabbene, 2013; Vidyasagar, 1997). A salient feature of this approach is the use of the theory of rare events and large deviation inequalities, which suitably bound the tail of the probability distribution. These inequalities are crucial in the area of statistical learning theory (Bousquet, Boucheron, & Lugosi, 2004; Vapnik, 1998), which has been utilized for feedback design of uncertain systems (Vidyasagar, 2001).

Design in the presence of uncertainty is of major relevance in different areas, including mathematical optimization and robustness (Ben-Tal & Nemirovski, 1998; Petersen & Tempo, 2014). The goal is to find a feasible solution which is optimal in some sense for all possible uncertainty instances. Unfortunately, the related semi-infinite optimization problems are often NP-hard (examples of NP-hard problems in systems and control can be found in Blondel & Tsitsiklis, 1997, 2000), and this may seriously limit their applicability from the computational point of view. There are two approaches to resolve this NP-hard issue. The first approach is based on the computation of deterministic relaxations of the original problem, which are usually polynomial time solvable. However, this might lead to overly conservative solutions (Scherer, 2006). An alternative is to assume that a probabilistic description of the uncertainty is available. Then, a randomized algorithm may be developed to compute, in polynomial time, a solution with probabilistic guarantees (Tempo et al., 2013; Vidyasagar, 1997). Stochastic programming methods (Prékopa, 1995) are similar in spirit to the methods studied in this paper and take advantage that, for random uncertainty, the underlying probability distributions are known or can be estimated. The goal is to find a solution that is feasible for almost all possible uncertainty realizations and maximizes the expectation of some function of the decisions variables.

☆ This work was supported by the MCYT-Spain under projects DPI2010-21589-C05-01, DPI2013-48243-C2-2-R and by the European Commission (FP7-257462). The work of Roberto Tempo was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 257462 HYCON2 Network of Excellence. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

E-mail addresses: alamo@cartuja.us.es (T. Alamo), roberto.tempo@polito.it (R. Tempo), amalia@cartuja.us.es (A. Luque), danirr@cartuja.us.es (D.R. Ramirez).

http://dx.doi.org/10.1016/j.automatica.2014.11.004
0005-1098/© 2014 Elsevier Ltd. All rights reserved.

The field of probabilistic methods (Calafiore, Dabbene, & Tempo, 2011; Tempo et al., 2013; Tempo & Ishii, 2007) has received growing attention in the systems and control community. Two complementary approaches, non-sequential and sequential, have been proposed. A classical approach for non-sequential methods is based upon statistical learning theory (Vapnik, 1998; Vidyasagar, 1997). Subsequent work along this direction includes Alamo, Tempo, and Camacho (2009), Chamanbaz, Dabbene, Tempo, Venkatakrishnan, and Wang (2014), Koltchinskii, Abdallah, Ariola, Dorato, and Panchenko (2000), Vidyasagar (2001) and Vidyasagar and Blondel (2001). Furthermore, in Alamo, Tempo, and Luque (2010a,b) and Luedtke and Ahmed (2008) the case in which the design parameter set has finite cardinality is analyzed. The advantage of these methods is that the problem under attention may be non-convex. For convex optimization problems, a non-sequential paradigm, denoted as the scenario approach, has been introduced in Calafiore and Campi (2005, 2006), see also Alamo et al. (2010b), Calafiore (2010) and Campi and Garatti (2008, 2011) for more advanced results, and Prandini, Garatti, and Vignali (2014) and Vayanos, Kuhn, and Rustem (2012) for recent developments in the areas of stochastic hybrid systems and multi-stage optimization, respectively. Finally, we refer to de Farias and Van Roy (2003) for a randomized approach to solve approximate dynamic programming.

In non-sequential methods, the original robustness problem is reformulated as a single optimization problem with sampled constraints, which are randomly generated. A relevant feature of these methods is that they do not require any validation step and the sample complexity is defined a priori. The main result of this line of research is to derive explicit lower bounds to the required sample size. However, the obtained explicit sample bounds can be overly conservative because they rely on a worst-case analysis and grow (at least linearly) with the number of decision variables.

For sequential methods, the resulting iterative algorithms are based on stochastic gradient (Calafiore & Polyak, 2001; Polyak & Tempo, 2001), ellipsoid iterations (Kanev, De Schutter, & Verhaegen, 2003; Oishi, 2007); or analytic center cutting plane methods (Calafiore & Dabbene, 2007; Wada & Fujisaki, 2009), see also Alamo, Tempo, Ramirez, and Camacho (2007) and Chamanbaz, Dabbene, Tempo, Venkataramanan, and Wang (2013) for other classes of sequential algorithms. Convergence properties in finite-time are one of the focal points of these papers. Various control problems have been solved using these sequential randomized algorithms, including robust LQ regulators (Polyak & Tempo, 2001), switched systems (Liberzon & Tempo, 2004) and uncertain linear matrix inequalities (LMIs) (Calafiore & Polyak, 2001). Sequential methods are often used for uncertain convex feasibility problems because the computational effort at each iteration is affordable. However, they have been studied also for non-convex problems, see Alamo et al. (2009) and Ishii, Basar, and Tempo (2005).

The common feature of most of these sequential algorithms is the use of the validation strategy presented in Dabbene, Shcherbakov, and Polyak (2010) and Oishi (2007). The candidate solutions provided at each iteration of these algorithms are tested using a validation set which is drawn according to the probability measure associated to the uncertainty. If the candidate solution satisfies the design specifications for every sampled element of this validation set, then it is classified as probabilistic solution and the algorithm terminates. The main point in this validation scheme is that the cardinality of the validation set increases very mildly at each iteration of the algorithm. The strategy guarantees that, if a probabilistic solution is obtained, then it meets some probabilistic specifications.

In this paper, we derive the sample complexity for various analysis and design problems related to uncertain systems. In particular we provide new results which guarantee that the tail of the binomial distribution is bounded by a pre-specified value. These results are then applied to the analysis of worst-case performance and constraint violation. With regard to design problems, we consider the special cases of finite families and robust convex optimization problems. This contribution improves the existing results by reducing the number of samples required to solve the design problem. We remark that the results we have obtained are fairly general and the assumptions on convexity and on finite families appear only in Section 4 which deals with probabilistic analysis and design.

The second main contribution of this paper is to propose a sequential validation scheme, denoted as Sequential Probabilistic Validation (SPV), which allows the candidate solution to violate the design specifications for one (or more) of the members of the validation set. The idea of allowing some violations of the constraints is not new and can be found, for example, in the context of system identification (Bai, Cho, Tempo, & Ye, 2002), chance-constrained optimization (Campi & Garatti, 2011; Nemirovski & Shapiro, 2006) and statistical learning theory (Alamo et al., 2009). This scheme makes sense in the presence of soft constraints or when a solution satisfying the specifications for all the admissible uncertainty realizations cannot be found. In this way, we improve the existing results with this relaxed validation scheme that reduces the chance of not detecting the solution even when it exists. Furthermore, we also show that a strict validation scheme may not be well-suited for some robust design problems.

This paper is based on the previous works of the authors Alamo, Luque, Ramirez, and Tempo (2012) and Alamo et al. (2010b). However, some results are completely new (Property 4) and others (Theorem 2, Property 1 and Property 3 and their proofs) are significant improvements of the preliminary results presented in the conference papers. Furthermore, the unifying approach studied here, which combines sample complexity results with SPV algorithms, was not present in previous papers. Finally, the numerical example in Section 8, which compares various approaches available in the literature, is also new. The rest of the paper is organized as follows. In the next section, we first introduce the problem formulation. In Section 3, we provide bounds for the binomial distribution which are used in Section 4 to analyze the probabilistic properties of different schemes involving randomization. In Section 5, we introduce the proposed family of probabilistically validated algorithms. The sample complexity of the validating sets is analyzed in Section 6. A detailed comparison with the validation scheme presented in Oishi (2007) is provided in Section 7. A numerical example where different schemes are used to address a robust identification problem is presented in Section 8. The paper ends with a section of conclusions and an Appendix which contains some auxiliary properties and proofs that are used in the previous sections.

2. Problem statement

We assume that a probability measure $\Pr_{\mathcal{W}}$ over the sample space $\mathcal{W}$ is given. Given $\mathcal{W}$, a collection of N independent identically distributed (i.i.d.) samples $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ drawn from $\mathcal{W}$ belongs to the Cartesian product $\mathcal{W}^N = \mathcal{W} \times \cdots \times \mathcal{W}$ (N times). Moreover, if the collection w of N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ is generated from $\mathcal{W}$ according to the probability measure $\Pr_{\mathcal{W}}$, then the *multisample* w is drawn according to the probability measure $\Pr_{\mathcal{W}^N}$. The scalars $\eta \in (0, 1)$ and $\delta \in (0, 1)$ denote probabilistic parameters called accuracy and confidence, respectively. Furthermore, $\ln(\cdot)$ is the natural logarithm and e is the Euler number. For $x \in \mathbb{R}$, $x \geq 0$, $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x; $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. For

$\alpha > 1$,

$$\xi(\alpha) := \sum_{k=1}^{\infty} \frac{1}{k^{\alpha}}$$

denotes the Riemann zeta function.

In a robustness problem, the controller parameters and auxiliary variables are parameterized by means of a decision variable vector θ, which is denoted as *design parameter* and is restricted to a set Θ. Furthermore, the uncertainty w is bounded in the set $\mathcal{W}$ and represents one of the admissible uncertainty realizations. We also consider a binary measurable function $g : \Theta \times \mathcal{W} \to \{0, 1\}$ and a real measurable function $f : \Theta \times \mathcal{W} \to \mathbb{R}$ which help to formulate the specific design problem under attention. More precisely, the binary function $g : \Theta \times \mathcal{W} \to \{0, 1\}$, is defined as

$$g(\theta, w) := \begin{cases} 0 & \text{if } \theta \text{ meets design specifications for } w \\ 1 & \text{otherwise,} \end{cases}$$

where design specifications are, for example, H_∞ norm bounds on the sensitivity function, see specific examples in Tempo et al. (2013), or the numerical example in Section 8.

Given $\theta \in \Theta$, the constraint $g(\theta, w) = 0$ is satisfied for a subset of $\mathcal{W}$. This concept is rigorously formalized by means of the notion of *probability of violation*, which is now introduced.

Definition 1 (*Probability of Violation*)**.** Consider a probability measure $\Pr_{\mathcal{W}}$ over $\mathcal{W}$ and let $\theta \in \Theta$ be given. The probability of violation of θ for the function $g : \Theta \times \mathcal{W} \to \{0, 1\}$ is defined as

$$E(\theta) := \Pr_{\mathcal{W}} \{g(\theta, w) = 1\}.$$

Using this notion we study the robust optimization problem

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } E(\theta) \le \eta, \tag{1}$$

where $J : \Theta \to (-\infty, \infty)$ is a measurable function which represents the controller performance and $\eta \in (0, 1)$ is a probabilistic accuracy. This is a chance constrained problem that may happen to be computationally intractable (Nemirovski & Shapiro, 2006). Given accuracy $\eta \in (0, 1)$ and confidence $\delta \in (0, 1)$, the main point of the probabilistic approach is to design an algorithm such that any probabilistic solution $\hat{\theta}$ obtained by running the algorithm, satisfies $E(\hat{\theta}) \le \eta$ with probability no smaller than $1 - \delta$.

Even in analysis problems when $\theta \in \Theta$ is given, it is often very hard to compute the exact value of the probability of violation $E(\theta)$ because this requires to solve a multiple integral with a usually non-convex domain of integration. However, we can approximate its value using the concept of empirical mean. For given $\theta \in \Theta$, and multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ drawn according to the probability measure $\Pr_{\mathcal{W}^N}$, the empirical mean of $g(\theta, w)$ with respect to w is defined as

$$\hat{E}(\theta, \mathrm{w}) := \frac{1}{N} \sum_{i=1}^{N} g(\theta, w^{(i)}).$$

Clearly, the empirical mean $\hat{E}(\theta, \mathrm{w})$ is a random variable. Since $g(\cdot, \cdot)$ is a binary function, $\hat{E}(\theta, \mathrm{w})$ is always within the closed interval $[0, 1]$.

The power of randomized algorithms stems from the fact that they can approximately solve non-convex design problems (with no-violation) of the type

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } g(\theta, w) = 0, \quad \text{for all } w \in \mathcal{W}. \tag{2}$$

In this setting, we draw N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$ and solve the sampled optimization problem

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } g(\theta, w^{(\ell)}) = 0, \quad \ell = 1, \ldots, N. \tag{3}$$

Since obtaining a global solution to this problem is still a difficult task in general, in this paper we analyze the probabilistic properties of *any suboptimal* solution. Furthermore, if at most m violations of the N constraints are allowed, the following sampled problem can be used to obtain a probabilistic relaxation to the original problem (2)

$$\min_{\theta \in \Theta} J(\theta) \text{ subject to } \sum_{\ell=1}^{N} g(\theta, w^{(\ell)}) \le m. \tag{4}$$

Randomized strategies to solve problems (3) and (4) have been studied in Alamo et al. (2009), see also Tempo et al. (2013). In order to analyze the probabilistic properties of any feasible solution to problem (4), we introduce the definitions of non-conforming feasible set and probability of failure.

Definition 2 (*Non-Conforming Feasible Set*)**.** Given N, the integer m where $0 \le m < N$, $\eta \in (0, 1)$, $g : \Theta \times \mathcal{W} \to \{0, 1\}$ and multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$, drawn according to the probability measure $\Pr_{\mathcal{W}^N}$, the non-conforming feasible set $\Theta(\mathrm{w}, \eta, m)$ is defined as

$$\Theta(\mathrm{w}, \eta, m) := \left\{\theta \in \Theta : \hat{E}(\theta, \mathrm{w}) \le \frac{m}{N} \text{ and } E(\theta) > \eta\right\}.$$

Definition 3 (*Probability of Failure*)**.** Given N, the integer m where $0 \le m < N$, $\eta \in (0, 1)$ and $g : \Theta \times \mathcal{W} \to \{0, 1\}$, the probability of failure, denoted by $p(N, \eta, m)$ is defined as

$$p(N, \eta, m) := \Pr_{\mathcal{W}^N}\{\Theta(\mathrm{w}, \eta, m) \text{ is not empty}\}.$$

The probability $p(N, \eta, m)$ defined here is slightly different than the probability of one-sided constrained failure introduced in Alamo et al. (2009). We notice that the non-conforming feasibility set is empty with probability $1 - p(N, \eta, m)$. This means that every feasible solution $\theta \in \Theta$ to problem (4) satisfies $E(\theta) \le \eta$ with probability $1 - p(N, \eta, m)$. Given the confidence parameter $\delta \in (0, 1)$, the objective is to obtain explicit expressions yielding a minimum number of samples N such that $p(N, \eta, m) \le \delta$.

3. Sample complexity for the binomial distribution

In this section, we provide bounds for the binomial distribution which are used in Section 4. Given a positive integer N and a nonnegative integer m, $m \le N$, and $\eta \in (0, 1)$, the binomial distribution function is given by

$$B(N, \eta, m) := \sum_{i=0}^{m} \binom{N}{i} \eta^i (1 - \eta)^{N-i}.$$

The problem we address in this section is the explicit computation of the *sample complexity*, i.e. a function $\tilde{N}(\eta, m, \delta)$ such that the inequality $B(N, \eta, m) \le \delta$ holds for any $N \ge \tilde{N}(\eta, m, \delta)$, where $\delta \in (0, 1)$. As illustrated in the following section, the inequality $B(N, \eta, m) \le \delta$ plays a fundamental role in probabilistic methods. Although some explicit expressions are available, e.g. the multiplicative and additive forms of Chernoff bound (Chernoff, 1952), the results obtained in this paper are tuned on the specific inequalities stemming from the problems described in Section 4.

The following technical lemma provides an upper bound for the binomial distribution $B(N, \eta, m)$.

Lemma 1. *Suppose that $\eta \in (0, 1)$ and that the nonnegative integer m and the positive integer N satisfy $m \le N$. Then, $B(N, \eta, m) \le a^m \left(\frac{\eta}{a} + 1 - \eta\right)^N$, $\forall a \ge 1$.*

Proof. The proof of the lemma follows from the following sequence of inequalities:

$$\begin{aligned} B(N,\eta,m) &= a^m \sum_{i=0}^{m} \binom{N}{i} a^{-m}\eta^i(1-\eta)^{N-i} \\ &\le a^m \sum_{i=0}^{m} \binom{N}{i} a^{-i}\eta^i(1-\eta)^{N-i} \\ &\le a^m \sum_{i=0}^{N} \binom{N}{i} \left(\frac{\eta}{a}\right)^i (1-\eta)^{N-i} \\ &= a^m \left(\frac{\eta}{a}+1-\eta\right)^N. \quad \square \end{aligned}$$

We notice that each particular choice of $a \ge 1$ provides an upper bound for $B(N,\eta,m)$. When using Lemma 1 to obtain a specific sample complexity, the selected value for a plays a significant role.

Lemma 2. *Given $\delta \in (0,1)$ and the nonnegative integer m, suppose that the integer N and the scalars $\eta \in (0,1)$ and $a > 1$ satisfy the inequality*

$$N \ge \frac{1}{\eta}\left(\frac{a}{a-1}\right)\left(\ln\frac{1}{\delta} + m\ln a\right). \tag{5}$$

Then, $m < N$ and $B(N,\eta,m) \le \delta$.

Proof. We first prove that if inequality (5) is satisfied, then $m < N$. Since $\eta \in (0,1)$ and $\delta \in (0,1)$, (5) implies

$$N > \left(\frac{a}{a-1}\ln a\right) m.$$

Next, we notice that

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{a}{a-1}\ln a\right) = \left(\frac{-1}{(a-1)^2}\right)\ln a + \frac{1}{a-1}.$$

Since $\ln a < a-1$ for every $a > 1$, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{a}{a-1}\ln a\right) > \left(\frac{-1}{(a-1)^2}\right)(a-1) + \frac{1}{a-1} = 0.$$

Using this fact, we conclude that $\frac{a}{(a-1)}\ln a$ is a strictly increasing function for $a > 1$. This means that

$$N > \left(\frac{a}{a-1}\ln a\right) m \ge \lim_{\hat{a}\to 1}\left(\frac{\hat{a}}{\hat{a}-1}\ln\hat{a}\right) m = m.$$

We now prove that (5) guarantees that $a^m(\frac{\eta}{a}+1-\eta)^N \le \delta$. The inequality (5) can be rewritten as

$$N\eta\left(\frac{a-1}{a}\right) \ge \ln\frac{1}{\delta} + m\ln a. \tag{6}$$

Since $x \le -\ln(1-x)$ for every $x \in (0,1)$, and $\eta(\frac{a-1}{a}) \in (0,1)$, from inequality (6), we obtain a sequence of inequalities

$$-N\ln\left(1-\eta\left(\frac{a-1}{a}\right)\right) \ge \ln\frac{1}{\delta} + m\ln a$$

$$\ln\delta \ge m\ln a + N\ln\left(1-\eta\left(\frac{a-1}{a}\right)\right)$$

$$\delta \ge a^m\left(\frac{\eta}{a}+1-\eta\right)^N.$$

We have therefore proved that inequality (5) implies $m \le N$ and $a^m(\frac{\eta}{a}+1-\eta)^N \le \delta$. The claim of the property follows directly from Lemma 1. $\square$

Obviously, the best sample size bound is obtained taking the infimum with respect to $a > 1$. However, this requires to solve numerically a one-dimensional optimization problem for given η, δ and m. We observe that a suboptimal value can be immediately obtained setting a equal to the Euler constant, which yields the sample complexity

$$N \ge \frac{1}{\eta}\left(\frac{\mathrm{e}}{\mathrm{e}-1}\right)\left(\ln\frac{1}{\delta} + m\right). \tag{7}$$

Since $\frac{\mathrm{e}}{\mathrm{e}-1} < 1.59$, we obtain $N \ge \frac{1.59}{\eta}\left(\ln\frac{1}{\delta}+m\right)$, which is (numerically) a significant improvement of the bound given in Calafiore (2010) and other bounds available in the literature (Calafiore et al., 2011). We also notice that, if $m > 0$, then the choice

$$a = 1 + \frac{\ln\frac{1}{\delta}}{m} + \sqrt{2\frac{\ln\frac{1}{\delta}}{m}}$$

provides a less conservative bound at the price of a more involved expression (Alamo et al., 2010b). Based on extensive numerical computations for several values of η, δ and m we conclude that this bound is very close to the "optimal" one. Note, however, that the optimal value can be obtained numerically using the Lambert W function (Corless, Gonnet, Hare, Jeffrey, & Knuth, 1996). In the next corollary, we present another more involved sample complexity bound which improves (7) for some values of the parameters.

Corollary 1. *Given $\delta \in (0,1)$ and the nonnegative integer m, suppose that the integer N and the scalar $\eta \in (0,1)$ satisfy the inequality*

$$N \ge \frac{1}{\eta}\left(m + \ln\frac{1}{\delta} + \sqrt{2m\ln\frac{1}{\delta}}\right). \tag{8}$$

Then, $m < N$ and $B(N,\eta,m) \le \delta$.

The proof of this corollary is shown in the Appendix.

4. Sample complexity for probabilistic analysis and design

We now study some problems in the context of randomized algorithms where one encounters inequalities of the form $B(N,\eta,m) \le \delta$. In particular, we show how the results of the previous section can be used to derive explicit sample size bounds which guarantee that the probabilistic solutions obtained from different randomized approaches meet some pre-specified probabilistic properties.

In Section 4.2 we derive bounds on $p(N,\eta,m)$ when Θ consists of a finite number of elements. On the other hand, if Θ consists of an infinite number of elements, a deeper analysis involving statistical learning theory is needed (Tempo et al., 2013; Vidyasagar, 1997). In Section 4.3 we study the probabilistic properties of the optimal solution of problem (3) under the assumption that $g(\theta,w) = 0$ is equivalent to $f(\theta,w) \le 0$, where $f : \Theta \times \mathcal{W} \to \mathbb{R}$ is a convex function with respect to θ in Θ. In this case, the result is not expressed in terms of probability of failure because it applies only to the optimal solution of problem (3), and not to every feasible solution.

4.1. Worst-case performance analysis

We recall a result shown in Tempo, Bai, and Dabbene (1997) for the probabilistic worst-case performance analysis.

Theorem 1. *Given the function $f : \Theta \times \mathcal{W} \to \mathbb{R}$ and $\hat{\theta} \in \Theta$, consider the multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ drawn from $\mathcal{W}^N$ according to probability $\Pr_{\mathcal{W}^N}$ and define $\gamma = \max_{\ell=1,\ldots,N} f(\hat{\theta}, w^{(\ell)})$. If*

$$N \geq \frac{\ln \frac{1}{\delta}}{\ln \frac{1}{1-\eta}},$$

then $\Pr_{\mathcal{W}}\{w \in \mathcal{W} : f(\hat{\theta}, w) > \gamma\} \leq \eta$ with probability no smaller than $1 - \delta$.

The proof of this statement can be found in Tempo et al. (1997) and is based on the fact that $\Pr_{\mathcal{W}}\{w \in \mathcal{W} : f(\hat{\theta}, w) > \gamma\} \leq \eta$ with probability no smaller than $1 - (1-\eta)^N$. Therefore, it suffices to take N such that $B(N, \eta, 0) = (1-\eta)^N \leq \delta$.

4.2. Finite families for design

We consider the non-convex sampled problem (4) for the special case when Θ consists of a set of finite cardinality n_C. As a motivation, we study the case when, after an appropriate normalization procedure, the design parameter set is rewritten as $\hat{\Theta} = \{\theta \in \mathbb{R}^{n_\theta} : \|\theta\|_\infty \leq 1\}$. Suppose also that a gridding approach is adopted. That is, for each component θ_j, $j = 1, \ldots, n_\theta$ of the design parameters $\theta \in \mathbb{R}^{n_\theta}$, only n_{C_j} equally spaced values are considered. That is, θ_j is constrained into the set $\Upsilon_j = \{-1 + \frac{2(t-1)}{(n_{C_j}-1)} : t = 1, \ldots, n_{C_j}\}$. With this gridding procedure, the following finite cardinality set $\Theta = \{[\theta_1, \ldots, \theta_{n_\theta}]^T : \theta_j \in \Upsilon_j,\ j = 1, \ldots, n_\theta\}$ is obtained. We notice that the cardinality of the set is $n_C = \prod_{j=1}^{n_\theta} n_{C_j}$. Another situation in which the finite cardinality assumption holds is when a finite number of random samples in the space of design parameter are drawn according to a given probability, see e.g. Fujisaki and Kozawa (2006) and Vidyasagar (2001).

The following theorem states the relation between the binomial distribution and the probability of failure under this finite cardinality assumption.

Theorem 2. *Suppose that the cardinality of Θ is no larger than n_C, $n_C > 0$, $\eta \in (0, 1)$ and $m < N$. Then,*

$$p(N, \eta, m) < n_C B(N, \eta, m).$$

Proof. If there is no element in Θ with probability of violation larger than η, then the non-conforming feasible set is empty for every multisample w and $p(N, \eta, m) = 0 < n_C B(N, \eta, m)$.

Suppose now that the subset of Θ of elements with probability of violation larger than η is not empty. Denote by $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(\tilde{n})}\}$ such a set. In this case, given a multisample w, the non-conforming feasible set is not empty if and only if the empirical mean is smaller than or equal to $\frac{m}{N}$ for at least one of the elements of this set. Therefore

$$\begin{aligned} p(N, \eta, m) &= \Pr_{\mathcal{W}^N}\{\Theta(\mathrm{w}, \eta, m) \text{ is not empty}\} \\ &= \Pr_{\mathcal{W}^N}\left\{\min_{1\leq k\leq \tilde{n}} \hat{E}(\theta^{(k)}, \mathrm{w}) \leq \frac{m}{N}\right\} \\ &\leq \sum_{k=1}^{\tilde{n}} \Pr_{\mathcal{W}^N}\left\{\hat{E}(\theta^{(k)}, \mathrm{w}) \leq \frac{m}{N}\right\} \\ &= \sum_{k=1}^{\tilde{n}} B(N, E(\theta^{(k)}), m) \\ &< \sum_{k=1}^{\tilde{n}} B(N, \eta, m) = \tilde{n} B(N, \eta, m). \end{aligned}$$

Notice that the last inequality is due to the fact that $E(\theta^{(k)}) > \eta$, $k = 1, \ldots, \tilde{n}$ and that the binomial distribution is a strictly decreasing function of η if $m < N$ (see Property 4 in the Appendix). To conclude the proof it suffices to notice that $\tilde{n} \leq n_C$. □

Consider now the optimization problem (4). It follows from Theorem 2 that to guarantee that every feasible solution $\hat{\theta} \in \Theta$ satisfies $E(\hat{\theta}) \leq \eta$ with probability no smaller than $1 - \delta$, it suffices to take $N > m$ such that $n_C B(N, \eta, m) \leq \delta$, where n_C is an upper bound on the cardinality of Θ. As it will be shown next, the required sample complexity in this case grows with the logarithm of n_C. This means that we can consider finite families with high cardinality and still obtain very reasonable sample complexity bounds.

Theorem 3. *Suppose that the cardinality of Θ is no larger than n_C. Given the nonnegative integer m, $\eta \in (0, 1)$ and $\delta \in (0, 1)$, if*

$$N \geq \inf_{a>1} \frac{1}{\eta}\left(\frac{a}{a-1}\right)\left(\ln \frac{n_C}{\delta} + m \ln a\right) \tag{9}$$

then $p(N, \eta, m) \leq \delta$. Moreover, if

$$N \geq \frac{1}{\eta}\left(m + \ln \frac{n_C}{\delta} + \sqrt{2m \ln \frac{n_C}{\delta}}\right)$$

then $p(N, \eta, m) \leq \delta$.

Proof. From Theorem 2 we have that $p(N, \eta, m) \leq \delta$ provided that $m < N$ and $B(N, \eta, m) \leq \frac{\delta}{n_C}$. The two claims of the property now follow directly from Lemma 2 and Corollary 1 respectively. □

From the definition of $p(N, \eta, m)$ and Theorem 3 we conclude that if one draws N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$, then with probability no smaller than $1 - \delta$, all the feasible solutions to problem (4) have a probability of violation no larger than η, provided that the cardinality of Θ is upper bounded by n_C and the sample complexity is given by

$$N \geq \frac{1}{\eta}\left(m + \ln \frac{n_C}{\delta} + \sqrt{2m \ln \frac{n_C}{\delta}}\right).$$

We remark that taking a equal to the Euler constant in (9), the following sample size bound

$$N \geq \frac{1}{\eta}\left(\frac{\mathrm{e}}{\mathrm{e}-1}\right)\left(\ln \frac{n_C}{\delta} + m\right)$$

is immediately obtained from Theorem 3. If $m > 0$, then a suboptimal value for a is given by

$$a = 1 + \frac{\ln \frac{n_C}{\delta}}{m} + \sqrt{2\frac{\ln \frac{n_C}{\delta}}{m}}.$$

4.3. Robust optimization for design

In this subsection, we study the so-called scenario approach for robust control introduced in Calafiore and Campi (2006). To address the semi-infinite optimization problem (2), we solve the randomized optimization problem (3). That is, we generate N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ from $\mathcal{W}$ according to the probability $\Pr_{\mathcal{W}}$ and then solve the following sampled optimization problem:

$$\min_{\theta\in\Theta} J(\theta) \text{ subject to } g(\theta, w^{(\ell)}) = 0, \quad \ell = 1, \ldots, N. \tag{10}$$

We consider here the particular case in which $J(\theta) = c^T\theta$, the constraint $g(\theta, w) = 0$ is convex in θ for all $w \in \mathcal{W}$ and the solution of (10) is unique. These assumptions are now stated precisely.

Assumption 1 (*Convexity*). Let $\Theta \subset \mathbb{R}^{n_\theta}$ be a convex and closed set. We assume that

$$J(\theta) := c^T\theta \quad \text{and} \quad g(\theta, w) := \begin{cases} 0 & \text{if } f(\theta, w) \le 0, \\ 1 & \text{otherwise} \end{cases}$$

where $f : \Theta \times \mathcal{W} \to [-\infty, \infty]$ is convex in θ for every fixed value of $w \in \mathcal{W}$.

Assumption 2 (*Feasibility and Uniqueness*). For all possible multisample extractions $\{w^{(1)}, \ldots, w^{(N)}\}$, the optimization problem (10) is always feasible and attains a unique optimal solution. Moreover, its feasibility domain has a nonempty interior.

Uniqueness may be assumed essentially without loss of generality, since in case of multiple optimal solutions one may always introduce a suitable tie-breaking rule (Calafiore & Campi, 2006). We now state a result that relates the binomial distribution to the probabilistic properties of the optimal solution obtained from (10), see Calafiore (2010) and Campi and Garatti (2008, 2011).

Lemma 3. *Let Assumptions 1 and 2 hold. Suppose that $N, \eta \in (0, 1)$ and $\delta \in (0, 1)$ satisfy the inequality*

$$\sum_{i=0}^{n_\theta - 1} \binom{N}{i} \eta^i (1-\eta)^{N-i} \le \delta. \tag{11}$$

Then, with probability no smaller than $1 - \delta$, the optimal solution $\hat{\theta}_N$ to the optimization problem (10) satisfies the inequality $E(\hat{\theta}_N) \le \eta$.

We now state an explicit sample size bound, which improves upon previous bounds, to guarantee that the probability of violation is smaller than η with probability at least $1 - \delta$.

Theorem 4. *Let Assumptions 1 and 2 hold. Given $\eta \in (0, 1)$ and $\delta \in (0, 1)$, if*

$$N \ge \inf_{a>1} \frac{1}{\eta}\left(\frac{a}{a-1}\right)\left(\ln\frac{1}{\delta} + (n_\theta - 1)\ln a\right) \tag{12}$$

or

$$N \ge \frac{1}{\eta}\left(\ln\frac{1}{\delta} + (n_\theta - 1) + \sqrt{2(n_\theta - 1)\ln\frac{1}{\delta}}\right) \tag{13}$$

then, with probability no smaller than $1 - \delta$, the optimal solution $\hat{\theta}_N$ to the optimization problem (10) satisfies the inequality $E(\hat{\theta}_N) \le \eta$.

Proof. From Lemma 3 it follows that it suffices to take N such that $B(N, \eta, n_\theta - 1) \le \delta$. Both inequalities (12) and (13) guarantee that $B(N, \eta, n_\theta - 1) \le \delta$ (see Lemma 2 and Corollary 1 respectively). This completes the proof. □

Taking a equal to the Euler constant in (12), we obtain

$$N \ge \frac{1}{\eta}\left(\frac{\mathrm{e}}{\mathrm{e}-1}\right)\left(\ln\frac{1}{\delta} + n_\theta - 1\right)$$

which improves the bound given in Calafiore (2010) and other bounds available in the literature (Calafiore et al., 2011). More precisely, the constant 2 appearing in Calafiore (2010) is reduced to $\frac{\mathrm{e}}{(\mathrm{e}-1)} \approx 1.59$, which is (numerically) a substantial improvement for small values of η. If $n_\theta > 1$, a suboptimal value for a is given by

$$a = 1 + \frac{\ln\frac{1}{\delta}}{n_\theta - 1} + \sqrt{2\frac{\ln\frac{1}{\delta}}{n_\theta - 1}}.$$

5. Sequential algorithms with probabilistic validation

In this section, we present a general family of randomized algorithms, which we denote as *Sequential Probabilistic Validation* (SPV) algorithms. The main feature of this class of algorithms is that they are based on a probabilistic validation step. This family includes most of the sequential randomized algorithms that have been presented in the literature and are discussed in the introduction of this paper.

Each iteration of an SPV algorithm includes the computation of a candidate solution for the problem and a subsequent validation step. The results provided in this paper are basically independent of the particular strategy chosen to obtain candidate solutions. Therefore, in the following discussion we restrict ourselves to a generic candidate solution. The accuracy $\eta \in (0, 1)$ and confidence $\delta \in (0, 1)$ required for the probabilistic solution play a relevant role when determining the sample size of each validation step. The main purpose of this part of the paper is to provide a validation scheme which guarantees that, for given accuracy η and confidence δ, all the probabilistic solutions obtained by running the SPV algorithm have a probability of violation no larger than η with probability no smaller than $1 - \delta$.

We enumerate each iteration of the algorithm by means of an integer k. We denote by m_k the number of violations that are allowed at the validation step of iteration k. We assume that m_k is a function of k, that is, $m_k = m(k)$ where the function $m : \mathbb{N} \to \mathbb{N}$ is given. We also denote by M_k the sample size of the validation step of iteration k. We assume that M_k is a function of k, η and δ. That is, $M_k = M(k, \eta, \delta)$ where $M : \mathbb{N} \times \mathbb{R} \times \mathbb{R} \to \mathbb{N}$ has to be appropriately designed in order to guarantee the probabilistic properties of the algorithm. In fact, one of the main contributions of Dabbene et al. (2010) and Oishi (2007) is to provide this function for the particular case $m_k = 0$ for every $k \ge 1$. The functions $m(\cdot)$ and $M(\cdot, \cdot, \cdot)$ are denoted as *level function* and *cardinality function* respectively.

We now introduce the structure of an SPV algorithm

(i) Set accuracy $\eta \in (0, 1)$ and confidence $\delta \in (0, 1)$ equal to the desired levels. Set k equal to 1.
(ii) Obtain a candidate solution $\hat{\theta}_k$ to the robust optimization problem (1).
(iii) Set $m_k = m(k)$ and $M_k = M(k, \eta, \delta)$.
(iv) Obtain validation set $\mathcal{V}_k = \{v^{(1)}, \ldots, v^{(M_k)}\}$ drawing M_k i.i.d. validation samples from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$.
(v) If $\sum_{\ell=1}^{M_k} g(\hat{\theta}_k, v^{(\ell)}) \le m_k$, then $\hat{\theta}_k$ is a probabilistic solution.
(vi) Exit if the exit condition is satisfied.
(vii) $k = k + 1$. Goto (ii).

Although the exit condition can be quite general, a reasonable choice is to exit after a given number of candidate solutions have been classified as probabilistic solutions or when a given computational time has elapsed since the starting of the algorithm. After exiting one could choose the probabilistic solution which maximizes a given performance index. We notice that in step (iv) we need to satisfy the i.i.d. assumption, and therefore sample reuse techniques are not applicable. In the next section, we propose a strategy to choose the cardinality of the validation set at iteration k in such a way that, with probability no smaller than $1 - \delta$, all candidate solutions classified as probabilistic solutions by the algorithm meet the accuracy η.

6. Adjusting the validation sample size

The cardinality adjusting strategy provided in this section constitutes a generalization of that presented in Dabbene et al.

(2010) and Oishi (2007). To obtain the results of this section we rely on some contributions on the sample complexity presented in the previous sections.

We now formally introduce the *failure function*.

Definition 4 (*Failure Function*)**.** The function $\mu : \mathbb{N} \to \mathbb{R}$ is said to be a failure function if it satisfies the following conditions:

(i) $\mu(k) \in (0, 1)$ for every positive integer k.
(ii) $\sum_{k=1}^{\infty} \mu(k) \leq 1$.

We notice that the function

$$\mu(k) = \frac{1}{\xi(\alpha)k^{\alpha}},$$

where $\xi(\cdot)$ is the Riemann zeta function, is a failure function for every $\alpha > 1$. This is due to the fact that $\sum_{i=1}^{\infty} \frac{1}{k^{\alpha}}$ converges for every scalar α greater than 1 to $\xi(\alpha)$. This family has been used in the context of validation schemes in Dabbene et al. (2010) and in Oishi (2007) for the particular value $\alpha = 2$.

Property 1. *Consider an SPV algorithm with given accuracy parameter $\eta \in (0, 1)$, confidence $\delta \in (0, 1)$, level function $m(\cdot)$ and cardinality function $M(\cdot, \cdot, \cdot)$. If $m(k) < M(k, \eta, \delta)$, for all $k \geq 1$, and there exists a failure function $\mu(\cdot)$ such that*

$$\sum_{i=0}^{m(k)} \binom{M(k, \eta, \delta)}{i} \eta^i (1-\eta)^{M(k,\eta,\delta)-i} \leq \delta\mu(k), \quad \forall k \geq 1$$

then, with probability greater than $1-\delta$, all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no greater than η.

The proof of this property follows the same lines as the proof of Theorem 9 in Oishi (2007).

Proof. We denote by δ_k the probability of classifying at iteration k the candidate solution $\hat{\theta}_k$ as a probabilistic solution under the assumption that the probability of violation $E(\hat{\theta}_k)$ is larger than η. Furthermore, let $M_k = M(k, \eta, \delta)$, then

$$\begin{aligned}\delta_k &= \Pr_{\mathcal{W}^{M_k}} \left\{ \hat{E}(\hat{\theta}_k, \mathrm{w}) \leq \frac{m_k}{M_k} \right\} \\ &= \sum_{i=0}^{m_k} \binom{M_k}{i} E(\hat{\theta}_k)^i (1-\hat{\theta}_k)^{M_k-i} \\ &< \sum_{i=0}^{m_k} \binom{M_k}{i} \eta^i (1-\eta)^{M_k-i}.\end{aligned}$$

Property 4 in the Appendix, $m_k < M_k$, and $E(\hat{\theta}_k) > \eta$ have been used to derive the last inequality. Then, we obtain

$$\delta_k < \sum_{i=0}^{m(k)} \binom{M(k, \eta, \delta)}{i} \eta^i (1-\eta)^{M(k,\eta,\delta)-i} \leq \delta\mu(k).$$

Therefore, the probability of misclassification of a candidate solution at iteration k is smaller than $\delta\mu(k)$. We conclude that the probability of erroneously classifying one or more candidate solutions as probabilistic solutions is bounded by

$$\sum_{k=1}^{\infty} \delta_k < \sum_{k=1}^{\infty} \delta\mu(k) = \delta \sum_{k=1}^{\infty} \mu(k) \leq \delta. \quad \square$$

To design a cardinality function $M(\cdot, \cdot, \cdot)$ satisfying the conditions of Property 1 we may use Corollary 1.

We now present the main contribution of this part of the paper, which is a general expression for the cardinality of the validation set at each iteration of the algorithm.

Theorem 5. *Consider an SPV algorithm with given accuracy $\eta \in (0, 1)$, confidence $\delta \in (0, 1)$ and level function $m(\cdot)$. Suppose also that $\mu(\cdot)$ is a failure function. Then, the cardinality function*

$$M(k, \eta, \delta) = \left\lceil \frac{1}{\eta} \left(m(k) + \ln \frac{1}{\delta\mu(k)} + \sqrt{2m(k) \ln \frac{1}{\delta\mu(k)}} \right) \right\rceil$$

guarantees that, with probability greater than $1-\delta$, all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no greater than η.

Proof. Corollary 1 guarantees that the proposed choice for the cardinality function satisfies $m(k) < M(k, \eta, \delta)$, for all $k \geq 1$, and

$$\sum_{i=0}^{m(k)} \binom{M(k, \eta, \delta)}{i} \eta^i (1-\eta)^{M(k,\eta,\delta)-i} \leq \delta\mu(k), \quad \forall k \geq 1.$$

The result then follows from a direct application of Property 1. $\square$

We notice that the proposed cardinality function $M(k, \eta, \delta)$ in Theorem 5 depends on the previous selection of the level function $m(\cdot)$ and the failure function $\mu(\cdot)$. A reasonable choice for these functions is $m(k) = \lfloor ak \rfloor$, where a is a non-negative scalar and $\mu(k) = \frac{1}{\xi(\alpha)k^{\alpha}}$ where α is greater than one. We recall that this choice guarantees that $\mu(k)$ is a failure function. As shown in the following section, the proposed level and failure functions allow us to recover, for the particular choice $a = 0$ the validation strategies proposed in Dabbene et al. (2010) and Oishi (2007). In the next corollary, we specify the generic structure of the SPV algorithm with the level function $m(k) = \lfloor ak \rfloor$, and state a probabilistic result.

Corollary 2. *Consider an SPV algorithm of the form given in Section 5 in which steps* (i) *and* (iii) *are substituted by*

(i) *Set accuracy $\eta \in (0, 1)$, confidence $\delta \in (0, 1)$ and scalars $a \geq 0$, $\alpha > 1$ equal to the desired levels. Set k equal to* 1.
(iii) *Set $m_k = \lfloor ak \rfloor$ and*

$$M_k = \left\lceil \frac{1}{\eta} \left(m_k + \ln \frac{\xi(\alpha)k^{\alpha}}{\delta} + \sqrt{2m_k \ln \frac{\xi(\alpha)k^{\alpha}}{\delta}} \right) \right\rceil.$$

Then, with probability greater than $1-\delta$, all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no greater than η.

Proof. The result is obtained directly from Theorem 5 using as level function $m(k) = \lfloor ak \rfloor$ and failure function $\mu(k) = \frac{1}{\xi(\alpha)k^{\alpha}}$. $\square$

Since the probabilistic properties of the algorithm presented in Corollary 2 are independent of the particular value of $\alpha > 1$, a reasonable choice for α is to select this parameter to minimize the cardinality of the validation sample set.

7. Comparison with other validation schemes

In this section, we provide comparisons with the validation schemes presented in Dabbene et al. (2010) and Oishi (2007). We notice that setting $a = 0$ and $\alpha = 2$ in Corollary 2 we obtain $m(k) = 0$ for every iteration k and

$$M(k) = \left\lceil \frac{1}{\eta} \ln \left(\frac{\xi(2)k^2}{\delta} \right) \right\rceil = \left\lceil \frac{1}{\eta} \ln \left(\frac{\pi^2 k^2}{6\delta} \right) \right\rceil.$$

This is the same cardinality function presented in Oishi (2007) if one takes into account that for small values of η, $-\ln(1-\eta)$ can be approximated by η. In the same way, $a = 0$ and $\alpha = 1.1$ lead to the cardinality function presented in Dabbene et al. (2010).

We notice that not allowing any failure in each validation test makes perfect sense for convex problems if the feasibility set $\Theta_r = \{\theta \in \Theta : g(\theta, w) = 0 \text{ for all } w \in \mathcal{W}\}$ is not empty. Under this assumption, the algorithm takes advantage of the validation samples that have not satisfied the specifications to obtain a new candidate solution. If Θ_r is not empty, a common feature of the methods which use this strict validation scheme is that a probabilistic solution (not necessarily belonging to the feasibility set Θ_r) is obtained in a finite number of iterations of the algorithm, see e.g., Alamo et al. (2007), Calafiore and Dabbene (2007) and Oishi (2007).

A very different situation is encountered when Θ_r is empty. We now state a property showing that a strict validation scheme ($a = 0$) should not be used to address the case of empty robust feasible set because the algorithm might fail to obtain a probabilistic solution even if the set $\{\theta \in \Theta : E(\theta) \le \eta\}$ is not empty.

Property 2. *Consider the SPV algorithm presented in Corollary 2 with $a = 0$ and $\alpha > 1$. Suppose that $E(\theta) \ge \mu > 0$ for all $\theta \in \Theta$. Then the SPV algorithm does not find any probabilistic solution in the first L iterations of the algorithm with probability greater than*

$$1 - \left(\frac{\delta}{\xi(\alpha)}\right)^{\frac{\mu}{\eta}} \Phi\left(\frac{\alpha\mu}{\eta}, \lceil \log_2 L \rceil\right),$$

where the function $\Phi(s, t)$ is given by

$$\Phi(s, t) := \begin{cases} \dfrac{1 - 2^{(1-s)(t+1)}}{1 - 2^{1-s}} & \text{if } s \neq 1 \\ t + 1 & \text{otherwise} \end{cases}$$

where s is a strictly positive scalar and t is a non-negative integer.

Proof. We notice that $a = 0$ implies that, at iteration k, the algorithm classifies a candidate solution $\hat{\theta}_k$ as a probabilistic solution only if it satisfies the constraint $g(\hat{\theta}_k, v^{(k)}) = 0,\ k = 1, \ldots, M_k$ where $\{v^{(1)}, \ldots, v^{(M_k)}\}$ is the randomly obtained validation set $\mathcal{V}_k$. Since $E(\theta) \ge \mu$ for all $\theta \in \Theta$ and $a = 0$, the probability of classifying a candidate solution as a probabilistic solution at iteration k is no greater than

$$\begin{aligned}(1-\mu)^{M_k} &= \mathrm{e}^{M_k \ln(1-\mu)} < \mathrm{e}^{-\mu M_k} \\ &\le \mathrm{e}^{-\frac{\mu}{\eta} \ln\left(\frac{\xi(\alpha)k^\alpha}{\delta}\right)} = \left(\frac{\delta}{\xi(\alpha)k^\alpha}\right)^{\frac{\mu}{\eta}}.\end{aligned}$$

Therefore, the probability of providing a probabilistic solution at any of the first L iterations of the algorithm is smaller than

$$\sum_{k=1}^{L} \left(\frac{\delta}{\xi(\alpha)k^\alpha}\right)^{\frac{\mu}{\eta}} = \left(\frac{\delta}{\xi(\alpha)}\right)^{\frac{\mu}{\eta}} \sum_{k=1}^{L} \left(\frac{1}{k^\alpha}\right)^{\frac{\mu}{\eta}}.$$

Taking $s = \frac{\alpha\mu}{\eta}$ and using Property 5 in the Appendix we have

$$\sum_{k=1}^{L} \left(\frac{1}{k^\alpha}\right)^{\frac{\mu}{\eta}} = \sum_{k=1}^{L} \frac{1}{k^s} \le \Phi(s, \lceil \log_2 L \rceil).$$

We conclude that the probability of not finding any probabilistic solution in the first L iterations of the algorithm is smaller than

$$1 - \left(\frac{\delta}{\xi(\alpha)}\right)^{\frac{\mu}{\eta}} \Phi\left(\frac{\alpha\mu}{\eta}, \lceil \log_2 L \rceil\right). \quad \square$$

We now present an example demonstrating that a strict validation scheme may not be well-suited for a robust design problem.

Example 1. Suppose that $\Theta = [0, 1]$, $\mathcal{W} = [-0.08, 1]$, $\eta = 0.1$, $\delta = 10^{-4}$ and

$$g(\theta, w) = \begin{cases} 0 & \text{if } \theta \le w \\ 1 & \text{otherwise.} \end{cases}$$

Suppose also that $\Pr_{\mathcal{W}}$ is the uniform distribution. It is clear that $\theta = 0$ minimizes the probability of violation and satisfies $\eta = 0.1 > E(0) = \frac{0.08}{1.08} > 0.074$. Therefore, we obtain $E(\theta) \ge 0.074 = \mu$ for all $\theta \in \Theta$. Consider now the choice $\alpha = 1.1$ and a maximum number of iterations L equal to 10^6. We conclude from Property 2 that, regardless of the strategy used to obtain candidate solutions, the choice $a = 0$ and $\alpha = 1.1$ in Corollary 2 will not provide a probabilistic solution with probability greater than 0.98. The choice $\alpha = 2$ leads to a probability greater than 0.99. This illustrates that a strict validation scheme is not well suited for this robust design problem. $\square$

The next result states that the probabilistic validation scheme presented in this paper achieves, under minor technical assumptions, a solution with probability one in a finite number of iterations.

Property 3. *Consider an SPV algorithm with given accuracy parameter $\eta \in (0, 1)$, confidence $\delta \in (0, 1)$ and level function $m(\cdot)$. Suppose that*

(i) *$\mu(\cdot)$ is a failure function.*
(ii) *The cardinality function $M(k, \eta, \delta)$ is given by*

$$\left\lceil \frac{1}{\eta}\left(m(k) + \ln\frac{1}{\delta\mu(k)} + \sqrt{2m(k)\ln\frac{1}{\delta\mu(k)}}\right)\right\rceil.$$

(iii) *There exist an integer k^*, scalars $\mu \in (0, 1)$ and $p \in (0, 1)$ such that at every iteration $k > k^*$ a candidate solution $\hat{\theta}_k$ satisfying $E(\hat{\theta}_k) \le \mu < \eta$ is obtained with probability greater than p.*
(iv) $\lim_{k\to\infty} \frac{1}{m(k)} \ln \frac{1}{\delta\mu(k)} = 0$.

Then, the SPV algorithm achieves with probability one a solution in a finite number of iterations.

Proof. Using the assumption

$$\lim_{k\to\infty} \frac{1}{m(k)} \ln \frac{1}{\delta\mu(k)} = 0$$

we conclude that

$$\begin{aligned}&\lim_{k\to\infty} \frac{M(k)}{m(k)} \\ &\quad = \lim_{k\to\infty} \frac{1}{\eta}\left(1 + \frac{1}{m(k)}\ln\frac{1}{\delta\mu(k)} + \sqrt{2\frac{1}{m(k)}\ln\frac{1}{\delta\mu(k)}}\right) = \frac{1}{\eta}.\end{aligned}$$

Since $\mu < \eta$ and $\frac{m(k)}{M(k)}$ converges to η, there exists $\tilde{k}$ such that

$$\frac{\mu + \eta}{2} \le \frac{m(k)}{M(k)}, \quad \text{for every } k > \tilde{k}.$$

That is, at each iteration k, the SPV algorithm provides candidate solutions $\hat{\theta}_k$ satisfying

$$E(\hat{\theta}_k) \le \mu \le \frac{\mu + \eta}{2} \le \frac{m(k)}{M(k)}$$

for every $k \ge \max\{k^*, \tilde{k}\}$ with probability greater than p. We notice that $\frac{1}{M(k)} \sum_{\ell=1}^{M(k)} g(\hat{\theta}_k, v^{(\ell)})$ is the empirical mean associated to $g(\hat{\theta}_k, v)$. We recall that the Chernoff inequality (see Tempo et al., 2013) guarantees that the probability of obtaining an empirical

mean larger than $\epsilon = \frac{\eta-\mu}{2}$ from the value $E(\hat{\theta}_k)$ is no larger than $e^{-2M(k)\epsilon^2}$. Notice that

$$E(\hat{\theta}_k) + \epsilon = E(\hat{\theta}_k) + \frac{\eta-\mu}{2} \leq \mu + \frac{\eta-\mu}{2} = \frac{\mu+\eta}{2} \leq \frac{m(k)}{M(k)}.$$

Therefore we have that if $k \geq \max\{k^*, \tilde{k}\}$, then with probability no smaller than $1 - e^{-2M(k)\epsilon^2}$ the candidate solution is classified as a probabilistic solution. Taking into account that $M(k)$ tends to infinity with k, there exists k_ϵ such that $1 - e^{-2M(k)\epsilon^2} \geq \frac{1}{2}$ for every $k > k_\epsilon$. This means that the probability of classifying a candidate solution as a probabilistic one is no smaller than $\frac{p}{2}$ for every iteration $k > \max\{k^*, \tilde{k}, k_\epsilon\}$. Since $\frac{p}{2} > 0$, we conclude that the algorithm obtains a probabilistic solution with probability one. □

8. Numerical example

The objective of this numerical example is to obtain probabilistic upper and lower bounds of a given time function $y : \mathcal{W} \to \mathbb{R}$ with unknown parameters A and B of the form

$$y(w) = \left[A\left(1 + \frac{1}{2}t^2\right)\sin(7t + 0.5) + B\right]e^{-\frac{3}{2}t},$$

where $w \in \mathcal{W}$. The uncertainty set $\mathcal{W}$ is

$$\mathcal{W} = \left\{w = \begin{bmatrix} t & A & B \end{bmatrix}^T,\ t \in [0, 1],\ A \in [1, 3],\ B \in [1, 3]\right\}.$$

For a given order d, we define the regressor $\varphi_d : \mathcal{W} \to \mathbb{R}^{d+1}$ as

$$\varphi_d(w) = \varphi_d\left(\begin{bmatrix} t & A & B \end{bmatrix}^T\right) = \begin{bmatrix} 1 & t & t^2 & \cdots & t^d \end{bmatrix}^T.$$

The objective of this example is to find a parameter vector $\theta = [\gamma_d, \lambda_d]^T$, $\gamma_d \in \mathbb{R}^{d+1}$ and $\lambda_d \in \mathbb{R}^{d+1}$ such that, with probability no smaller than $1 - \delta$,

$$\Pr_{\mathcal{W}}\{w \in \mathcal{W} : |y(w) - \gamma_d^T\varphi_d(w)| \geq \lambda_d^T|\varphi_d(w)|\} \leq \eta.$$

The vector $|\varphi_d(w)|$ is obtained from the absolute values of $\varphi_d(w)$. The binary function $g : \Theta \times \mathcal{W} \to \{0, 1\}$, is defined as

$$g(\theta, w) := \begin{cases} 0 & \text{if } \theta \text{ meets design specifications for } w \\ 1 & \text{otherwise,} \end{cases}$$

where "design specifications" means satisfying the following constraint:

$$|y(w) - \gamma_d^T\varphi_d(w)| \leq \lambda_d^T|\varphi_d(w)|$$

for uniformly randomly generated samples $w \in \mathcal{W}$.

A similar problem is addressed in Campi and Garatti (2008) using the scenario approach. For the numerical computations, we take $\delta = 10^{-6}$ and $\eta = 0.01$. We address the problem studying the finite families, scenario and SPV approach, and use the explicit sample complexity derived in the previous sections.

8.1. Finite families approach

We apply the results of Section 4.2 to determine both the degree d and the parameter vectors (γ_d, λ_d) that meet the design specification and optimize a given performance index.

In this example a finite family of cardinality $n_C = 400$ is considered. In order to compare the finite family approach with the scenario one, we consider no allowed failures (i.e. $m = 0$). For this choice of parameters ($m = 0$, $n_C = 400$, $\delta = 10^{-6}$ and $\eta = 0.01$), the number of samples N required to obtain a solution with the specified probabilistic probabilities is 1981 (see Theorem 3). A set $\mathcal{D}$ of $M = N$ samples is drawn (i.i.d.) from $\mathcal{W}$. We use these samples to select the optimal parameters $(\tilde{\gamma}_d, \tilde{\lambda}_d)$ corresponding to each of the different regressors $\varphi_d(\cdot)$. Each pair $(\tilde{\gamma}_d, \tilde{\lambda}_d)$ is obtained minimizing the empirical mean of the absolute value of the approximation error. That is, each pair $(\tilde{\gamma}_d, \tilde{\lambda}_d)$ is the solution to the optimization problem

$$\begin{aligned} \min_{\gamma_d, \lambda_d} \quad & \frac{1}{M}\sum_{w\in\mathcal{D}} \lambda_d^T|\varphi_d(w)| \\ \text{s.t.} \quad & |y(w) - \gamma_d^T\varphi_d(w)| \leq \lambda_d^T|\varphi_d(w)|, \quad \forall w \in \mathcal{D}. \end{aligned}$$

We notice that the obtained parameters do not necessarily satisfy the probabilistic design specifications. In order to resolve this problem, we consider a new set of candidate solutions of the form $\Theta = \{\theta_{d,j} = (\tilde{\gamma}_d, e^{\left(-0.5+\frac{j}{20}\right)}\tilde{\lambda}_d) : d = 1, \ldots, d_{\max}, j = 1, \ldots, j_{\max}\}$.

This family has cardinality $n_C = d_{\max}j_{\max}$. We take a large factor $e^{\left(-0.5+\frac{j}{20}\right)}$ to increase the probability to meet the design specifications. Therefore, choosing a large enough value for $j_{\max}$ leads to a non-empty intersection of Θ with the set of parameters that meet the design specifications. In this example, we take $j_{\max} = 20$ and $d_{\max} = 20$, which yields $n_C = 400$.

Using the finite family approach, we choose from Θ the design parameter that optimizes a given performance index. We draw from $\mathcal{W}$ a set $\mathcal{V}$ of N (i.i.d.) samples and select the pair that minimizes the empirical mean of the absolute value of the approximation error in the validation set $\mathcal{V}$. That is, we consider the performance index

$$\frac{1}{N}\sum_{w\in\mathcal{V}} e^{\left(-0.5+\frac{j}{20}\right)}\tilde{\lambda}_d^T|\varphi_d(w)|$$

subject to the constraints

$$|y(w) - \tilde{\gamma}_d^T\varphi_d(w)| \leq e^{\left(-0.5+\frac{j}{20}\right)}\tilde{\lambda}_d^T|\varphi_d(w)|, \quad \forall w \in \mathcal{V}.$$

We remark that the feasibility of this problem can be guaranteed in two ways. The first one is to choose $j_{\max}$ large enough. The second one is to allow m failures. As previously discussed, in this example we take $j_{\max} = 20$ and $m = 0$.

As the cardinality N of $\mathcal{V}$ has been chosen using Theorem 3, the probability of violation and the probability of failure of the best solution from Θ are bounded by η and δ respectively.

The obtained solution corresponds to $d = 15$ and $j = 11$ and the value for the performance index is 0.9814. Fig. 1 shows the approximation for the set $\mathcal{V}$ and the obtained probabilistic upper and lower bounds for the random function.

Finally, for illustrative purposes, we used a validation set of sample size $N_v = 10N$, obtaining 5 violations. The empirical violation probability turned out to be $\eta_{\exp} = \frac{5}{10N} = 2.5242 \cdot 10^{-4}$, while the specification was $\eta = 0.01$.

8.2. Convex scenario approach

In this case we take advantage of the result of Section 8.1 and choose $d = 15$ as the order of the approximation polynomial. Following the scenario approach we draw a set $\mathcal{W}_k$ of N samples (i.i.d.) from $\mathcal{W}$ and solve the convex optimization problem

$$\begin{aligned} \min_{\gamma_d, \lambda_d} \quad & \lambda_d^T E\{|\varphi_d(t)|\} \\ \text{s.t.} \quad & |y(w) - \gamma_d^T\varphi_d(w)| \leq \lambda_d^T|\varphi_d(w)|, \quad \forall w \in \mathcal{W}_k. \end{aligned}$$

In order to guarantee the design specifications we use Theorem 4 to determine the value of N. Since the number of decision variables is $2(d + 1) = 32$, $\eta = 0.01$ and $\delta = 10^{-6}$, the resulting value for N is 7090. We notice that the convex scenario approach

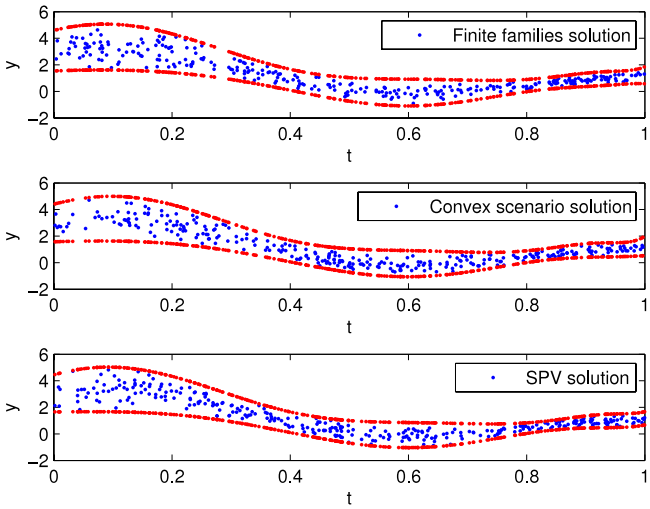

Fig. 1. Initial data set and envelope of the set of solutions.

does not apply directly to the minimization of the empirical mean. This is why one has to resort to the exact computation of the mean of the approximation error $\lambda_d^T E\{|\varphi_d(t)|\}$, see Campi and Garatti (2008). Fig. 1 shows the initial data set generated using the procedure described above, plus the envelope that contains all the polynomials.

For illustrative purposes, we check with a validation set of size $N_v = 10N$. The experimental value $\eta_{\text{exp}} = 8.4626 \cdot 10^{-5}$ is obtained, while the specification was $\eta = 0.05$. Using this strategy, 7090 samples are required, considerably bigger than in the finite families approach. We obtained a performance index of 0.9613, slightly better than that obtained by the finite families strategy. The advantage of the finite families approach is that, using a smaller number of samples, a similar performance is obtained. This allows us to determine the best order of the polynomial with the further advantage that the exact computation of the mean of the error is not required. Furthermore, the finite family approach does not rely on a convexity assumption.

8.3. SPV algorithm

We again take advantage of the result of Section 8.1 and choose $d = 15$ as the order of the approximation polynomial. Following the SPV algorithm approach, we begin setting $\eta = 0.01$, confidence $\delta = 10^{-6}$, scalars $a = 0.75$, $\alpha = 2$ and iteration index $k = 1$. The initial $\mathcal{W}_k$ is a set of 500 samples drawn from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$.

(i) A candidate solution $\hat{\theta}_k$ to the problem

$$\begin{aligned} \min_{\gamma_d, \lambda_d} \quad & \lambda_d^T \sum |\varphi_d(t)| \\ \text{s.t.} \quad & |y(w) - \gamma_d^T \varphi_d(w)| \le \lambda_d^T |\varphi_d(w)|, \quad \forall w \in \mathcal{W}_k \end{aligned}$$

is obtained.

(ii) Set $m_k = \lfloor ak \rfloor$ and

$$M_k = \left\lceil \frac{1}{\eta} \left(m_k + \ln \frac{\xi(\alpha) k^\alpha}{\delta} + \sqrt{2 m_k \ln \frac{\xi(\alpha) k^\alpha}{\delta}} \right) \right\rceil.$$

(iii) Obtain validation set $\mathcal{V}_k = \{v^{(1)}, \ldots, v^{(M_k)}\}$ drawing M_k i.i.d. validation samples from $\mathcal{W}$ according to the probability $\Pr_{\mathcal{W}}$.

(iv) If $\sum_{\ell=1}^{M_k} g(\hat{\theta}_k, v^{(\ell)}) \le m_k$, then $\hat{\theta}_k$ is a probabilistic solution.

(v) Exit if the exit condition is satisfied.

(vi) $k = k + 1$. $\mathcal{W}_k = \mathcal{W}_{k-1} \bigcup \mathcal{V}_{k-1}$. Goto (i).

Table 1
Sample complexity for different values of η.

η	N_{finite}	N_{convex}	N_{SPV}
0.1	398	488	988
0.05	794	849	1 972
0.01	3 962	7 090	4 163
0.005	7 924	16 078	13 652
0.001	39 614	74 062	41 617

Table 2
Performance index for different values of η.

η	J_{finite}	J_{convex}	J_{SPV}
0.1	1.0411	0.8803	0.8589
0.05	1.0085	0.9217	0.9597
0.01	0.9841	0.9613	0.9406
0.005	1.0111	1.0447	0.9741
0.001	0.9904	1.0183	0.9828

Fig. 1 shows the initial data set generated using the procedure described above, and the envelope that contains all the solution polynomials. Using this strategy, 4163 samples are required. We obtained a performance index of 0.9406, slightly better than the ones obtained by the other approaches.

The level function in the last step of the algorithm is $m_k = 1$, being the empirical probability of failure $\frac{1}{M_k} = \frac{1}{2231} < 0.01$.

Remark 1. If we set $a = 0$ in the algorithm, there are no allowed failures and this coincides with the approach studied in Oishi (2007). In this case, the algorithm did not find a solution for $\eta = 0.05$ and $M_k < 30\,000$. This is consistent with the results of Section 7.

In Table 1 the results of the three approaches are compared for different values of η. Note that N_{finite}, N_{convex} and N_{SPV} denote the total number of samples required in each of the three proposed strategies. We notice that, for small values of the probability of violation η, the sample complexity corresponding to the convex scenario is the largest one. On the other hand, as can be observed in Table 2, the performance index obtained with the SPV algorithms is slightly better than the ones corresponding to the other two approaches. We recall that the SPV algorithms do not rely on a convexity or finite cardinality assumptions.

9. Conclusions

In this paper, we have derived the *sample complexity* for various analysis and design problems related to uncertain systems. In particular, we provided new results which guarantee that a binomial distribution is smaller than a given probabilistic confidence. These results are subsequently exploited for analysis problems to derive the sample complexity of worst-case performance and robust optimization. With regard to design problems, these results can be used for finite families and for the special case when the design problem can be recast as a robust convex optimization problem.

We also presented a general class of randomized algorithms based on probabilistic validation, denoted as *Sequential Probabilistic Validation* (SPV). We provided a strategy to adjust the cardinality of the validation sets to guarantee that the obtained solutions meet the probabilistic specifications. The proposed strategy is compared with other existing schemes in the literature. In particular, it has been shown that a strict validation strategy where the design parameters need to satisfy the constraints for all the elements of the validation set might not be appropriate in some situations. We have shown that the proposed approach does not

suffer from this limitation because it allows the use of a non-strict validation test. As it has been shown in this paper, this relaxed scheme allows us to reduce, in some cases dramatically, the number of iterations required by the sequential algorithm. Another advantage of the proposed approach is that it does not rely on the existence of a robust feasible solution. Finally, we remark that this strategy is quite general and it is not based on finite families or convexity assumptions.

Appendix. Auxiliary proofs and properties

Proof of Corollary 1. We first notice that if $m = 0$, then $B(N, \eta, 0) = (1-\eta)^N = e^{N\ln(1-\eta)} \leq e^{-\eta N}$. Therefore, it follows from $\eta N \geq \ln\frac{1}{\delta}$ that $e^{-\eta N} \leq e^{\ln\delta} = \delta$. This proves the result for $m = 0$.

Consider now the case $m > 0$. We first prove that $h(r) := \sqrt{2(r-1)} - \ln\left(r + \sqrt{2(r-1)}\right) \geq 0$ for all $r \geq 1$. Since $h(1) = 0$, the inequality $h(r) \geq 0$ holds if the derivative of $h(r)$ is non negative for every $r > 1$.

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}r}h(r)\\
&= \frac{1}{\sqrt{2(r-1)}} - \frac{1}{r+\sqrt{2(r-1)}}\left(1 + \frac{1}{\sqrt{2(r-1)}}\right)\\
&= \left(\frac{1}{\sqrt{2(r-1)}}\right)\left(1 - \frac{1+\sqrt{2(r-1)}}{r+\sqrt{2(r-1)}}\right)\\
&= \left(\frac{1}{\sqrt{2(r-1)}}\right)\left(\frac{r-1}{r+\sqrt{2(r-1)}}\right) \geq 0, \quad \forall r > 1.
\end{aligned}$$

This proves the inequality $h(r) \geq 0$, for all $r \geq 1$. Denote now $\hat{a} = r + \sqrt{2(r-1)}$, with $r = 1 + \frac{1}{m}\ln\frac{1}{\delta}$. Clearly $\hat{a} > 1$. Therefore, from a direct application of Lemma 2, we conclude that it suffices to choose N such that

$$\begin{aligned}
N\eta &\geq \frac{\hat{a}}{\hat{a}-1}\left(\ln\frac{1}{\delta} + m\ln\hat{a}\right)\\
&= \frac{r+\sqrt{2(r-1)}}{r-1+\sqrt{2(r-1)}}\left(r - 1 + \ln\left(r+\sqrt{2(r-1)}\right)\right)m.
\end{aligned}$$

Since $h(r) \geq 0$ we conclude that

$$\frac{r-1+\ln\left(r+\sqrt{2(r-1)}\right)}{r-1+\sqrt{2(r-1)}} \leq 1.$$

From this inequality, we finally conclude that inequality $B(N, \eta, m) \leq \delta$ holds if

$$N\eta \geq (r+\sqrt{2(r-1)})m = m + \ln\frac{1}{\delta} + \sqrt{2m\ln\frac{1}{\delta}}. \quad \square$$

Property 4. *For fixed values of m and N, $m < N$, the binomial distribution function $B(N, \eta, m)$ is a strictly decreasing function of $\eta \in (0, 1)$.*

Proof. To prove the property, we show that the derivative of $B(N, \eta, m)$ with respect to η is negative. Let us define the scalars φ_i, $i = 0, \ldots, N$ as follows:

$$\begin{aligned}
\varphi_i(\eta) &:= \frac{\mathrm{d}}{\mathrm{d}\eta}\left(\eta^i(1-\eta)^{N-i}\right)\\
&= i\eta^{i-1}(1-\eta)^{N-i} - (N-i)\eta^i(1-\eta)^{N-i-1}\\
&= (i(1-\eta) - (N-i)\eta)\eta^{i-1}(1-\eta)^{N-i-1}\\
&= (i - N\eta)\eta^{i-1}(1-\eta)^{N-i-1}. \qquad \text{(A.1)}
\end{aligned}$$

With this definition we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\eta}B(N, \eta, m) &= \frac{\mathrm{d}}{\mathrm{d}\eta}\sum_{i=0}^{m}\binom{N}{i}\eta^i(1-\eta)^{N-i}\\
&= \sum_{i=0}^{m}\binom{N}{i}\varphi_i(\eta). \qquad \text{(A.2)}
\end{aligned}$$

We consider here two cases, $m - N\eta < 0$ and $m - N\eta \geq 0$. In the first case we have from Eq. (A.1) that $\varphi_i(\eta) < 0$, for $i = 0, \ldots, m$. This fact, along with Eq. (A.2) implies that the derivative with respect to η is negative and therefore the claim of the property is proved for this case.

Consider now the case $m - N\eta \geq 0$. In this case we have that $\varphi_i(\eta) > 0$, for $i > m$. Since $m < N$ we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\eta}B(N, \eta, m) &= \sum_{i=0}^{m}\binom{N}{i}\varphi_i(\eta)\\
&< \sum_{i=0}^{N}\binom{N}{i}\varphi_i(\eta)\\
&= \sum_{i=0}^{N}\binom{N}{i}\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\eta^i(1-\eta)^{N-i}\right)\\
&= \frac{\mathrm{d}}{\mathrm{d}\eta}\sum_{i=0}^{N}\binom{N}{i}\eta^i(1-\eta)^{N-i}\\
&= \frac{\mathrm{d}}{\mathrm{d}\eta}(\eta + (1-\eta))^N = \frac{\mathrm{d}}{\mathrm{d}\eta}(1)^N = 0.
\end{aligned}$$

We notice that in the last step of the proof the identity

$$(x+y)^N = \sum_{i=0}^{N}\binom{N}{i}x^iy^{N-i}$$

has been used. $\square$

Property 5. *Suppose that L is a positive integer and that s is a strictly positive scalar. Then, $\sum_{k=1}^{L}\frac{1}{k^s} \leq \Phi(s, \lceil\log_2 L\rceil)$ where, given $s \geq 0$ and the integer $t \geq 0$,*

$$\Phi(s, t) := \begin{cases} \dfrac{1-2^{(1-s)(t+1)}}{1-2^{1-s}} & \text{if } s \neq 1\\ t+1 & \text{otherwise.}\end{cases}$$

Proof. Given $L > 0$ and $s > 0$, define $t := \lceil\log_2(L)\rceil$ and $S(t) := \sum_{k=1}^{2^t}\frac{1}{k^s}$. Then we have $\sum_{k=1}^{L}\frac{1}{k^s} \leq \sum_{k=1}^{2^t}\frac{1}{k^s} = S(t)$. Next we show that $S(t) \leq 1 + 2^{1-s}S(t-1)$ for every integer t greater than 0. Since $S(0) = 1$ and $S(1) = 1 + 2^{-s}$, the inequality is clearly satisfied for $t = 1$. We now prove the inequality for t greater than 1

$$\begin{aligned}
S(t) &= \sum_{k=1}^{2^t}\frac{1}{k^s} = \sum_{k=1}^{2^{t-1}}\left[\frac{1}{(2k)^s} + \frac{1}{(2k-1)^s}\right]\\
&= 2^{-s}\sum_{k=1}^{2^{t-1}}\frac{1}{k^s} + \sum_{k=1}^{2^{t-1}}\frac{1}{(2k-1)^s}\\
&\leq 2^{-s}S(t-1) + 1 + \sum_{k=2}^{2^{t-1}}\frac{1}{(2k-2)^s}\\
&= 2^{-s}S(t-1) + 1 + 2^{-s}\sum_{k=1}^{2^{t-1}-1}\frac{1}{k^s}\\
&\leq 2^{-s}S(t-1) + 1 + 2^{-s}\sum_{k=1}^{2^{t-1}}\frac{1}{k^s}\\
&= 1 + 2^{1-s}S(t-1).
\end{aligned}$$

We have therefore proved the inequality $S(t) \leq 1+2^{1-s}S(t-1)$ for every integer t greater than 0. Using this inequality in a recursive way with $S(0) = 1$ we obtain $S(t) \leq \sum_{k=0}^{t} 2^{(1-s)k} = \Phi(s, t)$. This proves the result. □

References

Alamo, T., Luque, A., Ramirez, D. R., & Tempo, R. (2012). Randomized control design through probabilistic validation. In *Proceedings of the American control conference, Montreal, Canada*.

Alamo, T., Tempo, R., & Camacho, E. F. (2009). Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems. *IEEE Transactions on Automatic Control*, *54*(11), 2545–2559.

Alamo, T., Tempo, R., & Luque, A. (2010a). On the sample complexity of probabilistic analysis and design methods. In S. Hara, Y. Ohta, & J. C. Willems (Eds.), *Perspectives in mathematical system theory, control, and signal processing* (pp. 39–50). Berlin: Springer-Verlag.

Alamo, T., Tempo, R., & Luque, A. (2010b). On the sample complexity of randomized approaches to the analysis and design under uncertainty. In *Proceedings of the American control conference, Baltimore, USA*.

Alamo, T., Tempo, R., Ramirez, D. R., & Camacho, E. F. (2007). A sequentially optimal randomized algorithm for robust LMI feasibility problems. In *Proceedings of the European control conference, Kos, Greece*.

Bai, E. W., Cho, H., Tempo, R., & Ye, Y. (2002). Optimization with few violated constraints for linear bounded error parameter estimation. *IEEE Transactions on Automatic Control*, *47*(4), 1067–1077.

Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, *23*, 769–805.

Blondel, V., & Tsitsiklis, J. N. (1997). NP-hardness of some linear control design problems. *SIAM Journal on Control and Optimization*, *35*, 2118–2127.

Blondel, V., & Tsitsiklis, J. N. (2000). A survey of computational complexity results in system and control. *Automatica*, *36*, 1249–1274.

Bousquet, O., Boucheron, S., & Lugosi, G. (2004). Introduction to statistical learning theory. In O. Bousquet, U. Luxburg, & G. Raetsch (Eds.), *Advanced lectures on machine learning* (pp. 169–207). Berlin: Springer-Verlag.

Calafiore, G. (2010). Random convex programs. *SIAM Journal on Optimization*, *20*, 3427–3464.

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

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

Calafiore, G., & Dabbene, F. (2007). A probabilistic analytic center cutting plane method for feasibility of uncertain LMIs. *Automatica*, *43*, 2022–2033.

Calafiore, G., Dabbene, F., & Tempo, R. (2011). Research on probabilistic methods for control system design. *Automatica*, *47*, 1279–1293.

Calafiore, G., & Polyak, B. T. (2001). Stochastic algorithms for exact and approximate feasibility of robust LMIs. *IEEE Transactions on Automatic Control*, *11*(46), 1755–1759.

Campi, M. C., & Garatti, S. (2008). The exact feasibility of randomized solutions of robust convex programs. *SIAM Journal on Optimization*, *19*, 1211–1230.

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

Chamanbaz, M., Dabbene, F., Tempo, R., Venkatakrishnan, V., & Wang, Q.-G. (2014). A statistical learning theory approach for uncertain linear and bilinear matrix inequalities. *Automatica*, *50*, 1617–1625.

Chamanbaz, M., Dabbene, F., Tempo, R., Venkataramanan, V., & Wang, Q.-G. (2013). Sequential randomized algorithms for convex optimization in the presence of uncertainty. arXiv: arxiv.org/abs/1304.2222.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, *23*, 493–507.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., & Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics*, *5*, 329–359.

Dabbene, F., Shcherbakov, P. S., & Polyak, B. T. (2010). A randomized cutting plane method with probabilistic geometric convergence. *SIAM Journal on Optimization*, *20*, 3185–3207.

de Farias, D. P., & Van Roy, B. (2003). The linear programming approach to approximate dynamic programming. *Operations Research*, *51*, 850–865.

Fujisaki, Y., & Kozawa, Y. (2006). Probabilistic robust controller design: probable near minimax value and randomized algorithms. In G. Calafiore, & F. Dabbene (Eds.), *Probabilistic and randomized methods for design under uncertainty*. London: Springer.

Ishii, H., Basar, T., & Tempo, R. (2005). Randomized algorithms for synthesis of switching rules for multimodal systems. *IEEE Transactions on Automatic Control*, *50*, 754–767.

Kanev, S., De Schutter, B., & Verhaegen, M. (2003). An ellipsoid algorithm for probabilistic robust controller design. *Systems & Control Letters*, *49*, 365–375.

Koltchinskii, V., Abdallah, C. T., Ariola, M., Dorato, P., & Panchenko, D. (2000). Improved sample complexity estimates for statistical learning control of uncertain systems. *IEEE Transactions on Automatic Control*, *12*(45), 2383–2388.

Liberzon, D., & Tempo, R. (2004). Common Lyapunov functions and gradient algorithms. *IEEE Transactions on Automatic Control*, *49*, 990–994.

Luedtke, J., & Ahmed, S. (2008). A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, *2*(19), 674–699.

Nemirovski, A., & Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, *17*(4), 969–996.

Oishi, Y. (2007). Polynomial-time algorithms for probabilistic solutions of parameter-dependent linear matrix inequalities. *Automatica*, *43*, 538–545.

Petersen, I. R., & Tempo, R. (2014). Robust control of uncertain systems: classical results and recent developments. *Automatica*, *50*, 1315–1335.

Polyak, B. T., & Tempo, R. (2001). Probabilistic robust design with linear quadratic regulators. *Systems & Control Letters*, *43*, 343–353.

Prandini, M., Garatti, S., & Vignali, R. (2014). Performance assessment and design of abstracted models for stochastic hybrid systems through a randomized approach. *Automatica*, http://dx.doi.org/10.1016/j.automatica.2014.08.029.

Prékopa, A. (1995). *Stochastic programming*. Dordrecht: Kluwer.

Scherer, C. (2006). LMI relaxations in robust control. *European Journal of Control*, *12*, 3–29.

Tempo, R., Bai, E. W., & Dabbene, F. (1997). Probabilistic robustness analysis: explicit bounds for the minimum number of samples. *Systems & Control Letters*, *30*, 237–242.

Tempo, R., Calafiore, G., & Dabbene, F. (2013). *Randomized algorithms for analysis and control of uncertain systems, with applications* (2nd ed.). London: Springer-Verlag.

Tempo, R., & Ishii, H. (2007). Monte Carlo and Las Vegas randomized algorithms for systems and control: an introduction. *European Journal of Control*, *13*, 189–203.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: John Wiley and Sons.

Vayanos, P., Kuhn, D., & Rustem, B. (2012). A constraint sampling approach for multi-stage robust optimization. *Automatica*, *48*, 459–471.

Vidyasagar, M. (1997). *A theory of learning and generalization: with applications to neural networks and control systems*. London: Springer.

Vidyasagar, M. (2001). Randomized algorithms for robust controller synthesis using statistical learning theory. *Automatica*, *37*, 1515–1528.

Vidyasagar, M., & Blondel, V. D. (2001). Probabilistic solutions to some NP-hard matrix problems. *Automatica*, *37*, 1397–1405.

Wada, T., & Fujisaki, Y. (2009). Probabilistic cutting plane technique based on maximum volume ellipsoid center. In *Proceedings of the IEEE conference decision and control and the Chinese control conference, Shanghai, China* (pp. 1169–1174).

Teodoro Alamo was born in Spain in 1968. He received his M.Eng. degree in Telecommunications Engineering from the Polytechnic University of Madrid (Spain) in 1993 and his Ph.D. in Telecommunications Engineering from the University of Seville in 1998. From 1993 to 2000 he was an assistant professor of the Department of System Engineering and Automatic Control of the University of Seville. He was an associate professor from 2001 to 2010. Since March 2010 he is full professor in the same department. He was at the École Nationale Supérieure des Télécommunications (Telecom Paris) from September 1991 to May 1993. Part of his Ph.D. was done at RWTH Aachen, Alemania, June–September 1995. He is the author or coauthor of more than 160 publications including books, book chapters, journal papers, conference proceedings and educational books (google scholar profile available at http://scholar.google.es/citations?user=W3ZDTkIAAAAJ&hl=en). His current research interests are in decision making, model predictive control, machine learning, randomized algorithms and optimization strategies. He has co-funded the spin off company Optimal Performance (University of Seville, Spain).

Roberto Tempo is currently a Director of Research of Systems and Computer Engineering at CNR-IEIIT, Politecnico di Torino, Italy. He has held visiting positions at Chinese Academy of Sciences in Beijing, Kyoto University, The University of Tokyo, University of Illinois at Urbana-Champaign, German Aerospace Research Organization in Oberpfaffenhofen and Columbia University in New York. His research activities are focused on the analysis and design of complex systems with uncertainty, and various applications within information technology. On these topics, he has published more than 180 research papers in international journals, books and conferences. He is also a co-author of the book Randomized Algorithms for Analysis and Control of Uncertain Systems, with Applications, Springer-Verlag, London, published in two editions in 2005 and 2013.

He is a Fellow of the IEEE and a Fellow of the IFAC. He is a recipient of the IFAC Outstanding Paper Prize Award for a paper published in Automatica and of the Distinguished Member Award from the IEEE Control Systems Society. He is a Corresponding Member of the Academy of Sciences, Institute of Bologna, Italy, Class Physical Sciences, Section Technical Sciences.

In 2010 Dr. Tempo was President of the IEEE Control Systems Society. He is now serving as Editor-in-Chief of Automatica. He has been Editor for Technical Notes and Correspondence of the IEEE Transactions on Automatic Control in 2005–2009 and a Senior Editor of the same journal in 2011–2014. He is a member of the Advisory Board of Systems & Control: Foundations & Applications, Birkhauser. He was General Co-Chair for the IEEE Conference on Decision and Control, Florence, Italy, 2013 and Program Chair of the first joint IEEE Conference on Decision and Control and European Control Conference, Seville, Spain, 2005.

Amalia Luque has Ph.D. in Automation and Robotics from the University of Seville, 2014, where she works at present and M.S. in Automation, Robotics and Telematics from University of Seville. Her research interests include randomized algorithms, renewable energies and software development.

Daniel R. Ramirez was born in Spain in 1972. He received M.Eng. and Ph.D. degrees in Computer Engineering from University of Seville (Spain) in 1996 and 2002, respectively. From 1997 to 1999 he was a research assistant of the Department of System Engineering and Automatic Control. Since 1999, he has been an Assistant Professor and Associate Professor in the same department. He has authored and co-authored more than 50 technical papers in international journals and conference proceedings. He has participated in several European Projects and his current research interests include randomized algorithms, model predictive control and soft computing techniques.