

# A Probabilistic Approach to Optimal Estimation – Part II: Algorithms and Applications

Fabrizio Dabbene, Mario Sznaier, and Roberto Tempo

**Abstract**—In this paper, we develop randomized and deterministic algorithms for computing the probabilistic radius of information associated to an identification problem, and the corresponding optimal probabilistic estimate. To compute this estimate, in the companion paper [11] the concept of *optimal violation function* is introduced. Moreover, for the case of uniform distributions, it is shown how its computation is related to the solution of a (quasi) concave optimization problem, based on to the maximization of the volume of a specially constructed polytope. In this second paper, we move a step further and develop specific algorithms for addressing this problem. In particular, since the problem is NP-hard, we propose both randomized relaxations (based on a probabilistic volume oracle and stochastic optimization algorithms), and deterministic ones (based on semi-definite programming). Finally, we present a numerical example illustrating the performance of the proposed algorithms.

**Keywords:** System identification, optimal algorithms, randomized algorithms, uncertain systems

## I. INTRODUCTION

In the companion paper [11], a probabilistic framework has been introduced in dealing with system identification in the presence of bounded errors for which statistical information is available. The proposed approach provides a rigorous rapprochement between the classical paradigm, based on statistical assumptions on the error, and the set-membership approach, based on a purely deterministic description of the error. The main idea in this rapprochement is to discard sets of measure at most  $\epsilon$ , where  $\epsilon$  is a probabilistic accuracy, from the set of deterministic estimates, thus decreasing the so-called worst-case radius of information at the expense of a given probabilistic risk. This approach may be embedded in the probabilistic setting of information-based complexity (IBC), a theoretical framework developed within computer science, see e.g. [25].

The main theoretical results have been established in [11]; here we concentrate instead in developing computationally efficient algorithms for obtaining probabilistic estimates. In particular, we introduce specific algorithms for computing the *optimal violation function*  $v_o(r)$  for the case when the uncertainty is uniformly distributed. First, in Section III we observe that the exact computation of  $v_o(r)$  requires the evaluation of the volume of polytopes, which could be performed only for small dimensions. Since this problem is NP-Hard [17], in Section IV we propose to use suitable probabilistic

relaxations. More precisely, we present a randomized algorithm based upon the classical Markov Chain Monte Carlo method [23], [22], which has been studied in the context of randomized stochastic approximation methods [18]; see also [24] and [7] for further details about randomized algorithms. Then, in Section V, we present a deterministic relaxation of  $v_o(r)$  which is based upon the solution of a semi-definite program (SDP). The performance of both algorithms is tested using the example of identification of a FIR(3) system presented in Section VI.

## A. Notation

In this section, we provide the notation used in this paper. We write  $\|\cdot\|$ ,  $\|\cdot\|_2$  and  $\|\cdot\|_\infty$  to denote the  $\ell_p$ ,  $\ell_2$  and  $\ell_\infty$  norms, respectively. The  $\ell_p$  norm-ball of center  $\xi_c$  and radius  $r$  is denoted by  $\mathcal{B}(\xi_c, r) \doteq \{\xi \mid \|\xi - \xi_c\| \leq r\}$ , and we write  $\mathcal{B}(r) \doteq \mathcal{B}(0, r)$ . We denote by  $\mathcal{B}_2(\xi_c, r)$  and by  $\mathcal{B}_\infty(\xi_c, r)$  the  $\ell_2$  and  $\ell_\infty$  norm-balls, respectively. We use the notation  $x \sim p_A$  to indicate that the random vector  $x$  has probability density function (pdf)  $p_A(x)$  with support set  $A$ . The *uniform density*  $\mathcal{U}_A$  over the set  $A \subset \mathbb{R}^n$  is defined as

$$\mathcal{U}_A(x) \doteq \begin{cases} 1/\text{vol}[A] & \text{if } x \in A; \\ 0 & \text{otherwise} \end{cases},$$

where  $\text{vol}[A]$  represents the Lebesgue measure (volume) of the set  $A$ , see [15] for details regarding volumes of sets. The (univariate) *unilateral Gamma density* with parameters  $a, b \in \mathbb{R}$  is defined as  $G_{a,b}(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b}$ ,  $x \geq 0$ , where  $\Gamma(\cdot)$  is the Gamma function. We denote by  $\mathbb{I}(\cdot)$  the indicator function, which is equal to one if the clause is true, and it is zero otherwise. The  $n \times n$  identity matrix is indicated by  $I_n$ .

## II. SUMMARY OF THE IBC SETTING

In [11] we formally introduced the information-based complexity framework and its specific connections to system identification. In this section, for completeness and self-consistency, we provide a brief summary of the IBC setup. The starting point is a set of (unknown) problem elements  $x \in X \subseteq \mathbb{R}^n$ , of which corrupted information, or *data*,  $y \in Y \subseteq \mathbb{R}^m$  is available, i.e.

$$y = \mathcal{I}x + q \quad (1)$$

where  $\mathcal{I}$  is a *linear* information operator, and  $q$  represents *uncertainty*. In particular, in this paper we assume that  $q \in \mathcal{Q}$ , where  $\mathcal{Q} \equiv \mathcal{B}(\rho) \subseteq \mathbb{R}^m$  is an  $\ell_p$ -norm ball. Moreover,

F. Dabbene and R. Tempo are with the CNR-IEIT Institute, Politecnico di Torino, Italy (e-mail: fabrizio.dabbene@polito.it, roberto.tempo@polito.it).

M. Sznaier is with the Northeastern University, Boston, MA 02115, USA (e-mail: msznaier@ece.neu.edu).

we assume that  $q$  is uniformly distributed over  $\mathcal{Q}$ , that is  $q \sim \mathcal{U}_{\mathcal{Q}}$ .

Secondly, the solution operator  $\mathcal{S} : X \rightarrow Z$  is introduced, where  $Z \subseteq \mathbb{R}^s$  with  $s \leq n$ . Given  $\mathcal{S}$ , our aim is to estimate an element  $\mathcal{S}x \in Z$  knowing the corrupted information  $y \in Y$  about the problem element  $x \in X$ . That is, we construct an *algorithm*  $\mathcal{A} : Y \rightarrow Z$  mapping the data  $y$  into an (approximate) solution  $z = \mathcal{A}(y)$ , called an *estimator*.

Next, we define the so-called *consistency set*

$$\mathcal{I}^{-1}(y) \doteq \{x \in X \mid \text{there exists } q \in \mathcal{Q} : y = \mathcal{I}x + q\} \quad (2)$$

representing the set of all problem elements  $x \in X$  compatible with (i.e. not invalidated by)  $\mathcal{I}x$ , uncertainty  $q$  and bounding set  $\mathcal{Q}$ . Similarly, the set  $\mathcal{S}\mathcal{I}^{-1}(y)$  is defined as the mapping of  $\mathcal{I}^{-1}(y)$  through the solution operator. The following assumptions are made in [11] without loss of generality, see this paper for further discussions.

*Assumption 1 (Sufficient information and feasibility):*

We assume that the information operator  $\mathcal{I}$  is a one-to-one mapping, i.e.  $m \geq n$  and  $\text{rank } \mathcal{I} = n$ . Similarly,  $n \geq s$  and  $\mathcal{S}$  is full row rank. Moreover, we assume that the set  $\mathcal{I}^{-1}(y)$  has non-empty interior.

*Assumption 2 (Regularized solution operator):* In the sequel, we assume that the solution operator is regularized, so that  $\mathcal{S} = [\bar{\mathcal{S}} \ 0_{s, n-s}]$ , with  $\bar{\mathcal{S}} \in \mathbb{R}^{s, s}$ .

#### A. Optimal algorithms and violation functions

Based on these definitions, in the classical set-membership setup, given corrupted information  $y \in Y$ , a worst-case optimal algorithm minimizes the maximum distance between the estimate and the true-but-unknown solution. In [11], an alternative probabilistic notion is introduced. Namely, given an accuracy level  $\epsilon \in (0, 1)$ , we define the probabilistic error (to level  $\epsilon$ )  $r^{\text{Pr}}(\mathcal{A}, y, \epsilon)$  of the algorithm  $\mathcal{A}$  as

$$r^{\text{Pr}}(\mathcal{A}, y, \epsilon) \doteq \inf_{\mathcal{X}_\epsilon \text{ such that } \bar{\mu}_{\mathcal{I}^{-1}}(\mathcal{X}_\epsilon) \leq \epsilon} \max_{x \in \mathcal{I}^{-1}(y) \setminus \mathcal{X}_\epsilon} \|\mathcal{S}x - \mathcal{A}(y)\| \quad (3)$$

where the notation  $\mathcal{I}^{-1}(y) \setminus \mathcal{X}_\epsilon$  indicates the set-theoretic difference between  $\mathcal{I}^{-1}(y)$  and  $\mathcal{X}_\epsilon$ . Then, a *probabilistic optimal* (to level  $\epsilon$ ) algorithm  $\mathcal{A}_\epsilon^{\text{Pr}}$  is the one that minimizes the error  $r^{\text{Pr}}(\mathcal{A}, y, \epsilon)$  for any  $y \in Y$  and  $\epsilon \in (0, 1)$ . That is, given data  $y \in Y$  and accuracy level  $\epsilon \in (0, 1)$ , we have

$$r_\epsilon^{\text{Pr}}(y, \epsilon) \doteq r^{\text{Pr}}(\mathcal{A}_\epsilon^{\text{Pr}}, y, \epsilon) = \inf_{\mathcal{A}} r^{\text{Pr}}(\mathcal{A}, y, \epsilon). \quad (4)$$

The minimal error  $r_\epsilon^{\text{Pr}}(y, \epsilon)$  is called the *probabilistic radius of information* (to level  $\epsilon$ ) and the corresponding optimal estimator is given by

$$z_\epsilon^{\text{Pr}}(\epsilon) \doteq \mathcal{A}_\epsilon^{\text{Pr}}(y, \epsilon). \quad (5)$$

In [11] we analyze the problem of computing the probabilist optimal radius of information, and the corresponding probabilistic optimal estimate. In particular, we show that  $r_\epsilon^{\text{Pr}}(y, \epsilon)$  is the solution of the following chance-constrained optimization problem

$$r_\epsilon^{\text{Pr}}(y, \epsilon) = \min \{r \mid v_o(r) \leq \epsilon\}, \quad (6)$$

where the *optimal violation function* for a given radius  $r$  is defined as

$$v_o(r) \doteq \inf_{\mathcal{A}} \bar{\mu}_{\mathcal{I}^{-1}} \{x \in \mathcal{I}^{-1}(y) : \|\mathcal{S}x - \mathcal{A}(y)\| > r\}. \quad (7)$$

Note that the optimal violation function  $v_o(r)$  plays a key role in our setup. Indeed, in [11, Theorem 2] we prove that, for  $r > 0$ ,  $v_o(r)$  is right-continuous and non-increasing. Moreover, in the same theorem we show that  $v_o(r)$  can be computed by maximizing a specially constructed volume function  $\phi(z, r)$ . More precisely, for given center  $z \in Z$  and radius  $r > 0$ , we introduce the cylinder in  $X$  as follows

$$\mathcal{C}(z, r) \doteq \{x \in \mathbb{R}^n \mid \|\mathcal{S}x - z\| \leq r\} \subset X, \quad (8)$$

and the corresponding intersection set between  $\mathcal{C}(z, r)$  and the consistency set  $\mathcal{I}^{-1}(y)$

$$\Phi(z, r) \doteq \mathcal{I}^{-1}(y) \cap \mathcal{C}(z, r) \subset X. \quad (9)$$

Then, the function  $\phi(z, r)$  is defined as the volume  $\Phi(z, r)$ , i.e.

$$\phi(z, r) \doteq \text{vol}[\Phi(z, r)]. \quad (10)$$

Then, [11, Theorem 2] states that

$$v_o(r) = 1 - \frac{\phi_o(r)}{\text{vol}[\mathcal{I}^{-1}(y)]}, \quad (11)$$

where  $\phi_o(r)$  is the solution of the optimization problem

$$\text{(P-max-int)} : \max_{z \in \mathcal{H}(r)} \phi(z, r) \quad (12)$$

and  $\mathcal{H}(r)$  is the set of all centers  $z \in \mathbb{R}^s$  for which the intersection set  $\Phi(z, r)$  is non-empty.

The aim of this paper is to discuss different algorithmic approaches for the solution of problem (P-max-int). First, note that this problem is very hard in general. For instance, for  $\ell_1$  or  $\ell_\infty$  norms, the consistency set  $\mathcal{I}^{-1}(y)$  is a polytope and  $\mathcal{C}(z, r)$  is a cylinder parallel to the coordinate axes whose cross-section is a polytope. Hence, even evaluating the function  $\phi(z, r)$  appearing in (12) amounts to computing the volume of a polytope, and this problem has been shown to be NP-hard in [17]. This is further discussed in the next section, which deals with “exact” computational solutions.

### III. VOLUME ORACLE AND ORACLE-POLYNOMIAL-TIME ALGORITHM

For the case of polytopic sets, the papers [1], [14] study the problem (P-max-int) in the hypothetical setting that an oracle exists which satisfies the following property: given  $r > 0$  and  $z \in \mathcal{H}(r)$ , it returns the value of the function  $\phi(z, r)$ , together with a sub-gradient of it. In this case, in [1] a strongly polynomial-time (*in the number of oracle calls*) algorithm is derived. Note that, even if the problem is NP-hard in general, one can compute the volume of a polytope in a reasonable time for considerably complex polytopes in modest (e.g. for  $n \leq 10$ ) dimensions, see [4]. In this particular case, for  $\ell_\infty$  norms, the method proposed by [14] may be used. For instance, in the example discussed in Section VI, all relevant quantities have been computed exactly by

employing this method. However, it should be remarked that, for larger dimensions, the curse of dimensionality makes the problem computationally intractable, and alternative methods need to be devised.

In the next sections, we develop randomized and deterministic relaxations of problem (P-max-int) which do not suffer from these computational drawbacks.

#### IV. RANDOMIZED ALGORITHMS FOR COMPUTING (P-MAX-INT)

In this section, we propose randomized algorithms based on a probabilistic volume oracle and a stochastic optimization approach for approximately solving problem (P-max-int) for generic  $\ell_p$  norms. First of all, we compute a bounded version of the cylinder  $\mathcal{C}(z, r)$ . To this end, we note that bounds  $x_i^-, x_i^+$  on the variables  $x_i, i = s+1, \dots, n$ , can be computed as the solution of the following  $2(n-s)$  convex programs,

$$x_i^- = \min x_i \quad \text{subject to } x \in \mathcal{I}^{-1}(y) \quad (13)$$

$$x_i^+ = \max x_i \quad \text{subject to } x \in \mathcal{I}^{-1}(y) \quad (14)$$

for  $i = s+1, \dots, n$ . These problems are convex, and for generic  $\ell_p$  norms can be solved by any gradient-based method. In particular, problem (13)-(14) reduces to the solution to  $2(n-s)$  linear programs in the case of  $\ell_1$  or  $\ell_\infty$  norms. Then, under Assumption 2, we define the cylinder

$$\bar{\mathcal{C}}(z, r) \doteq \left\{ x \in \mathbb{R}^n \mid \left\| \bar{\mathcal{S}} \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} - z \right\| \leq r, \right. \\ \left. x_i^- \leq x_i \leq x_i^+, i = s+1, \dots, n \right\}.$$

Note that the cylinder  $\bar{\mathcal{C}}(z, r)$  is bounded, and has volume equal to

$$\text{vol}[\bar{\mathcal{C}}(z, r)] = \frac{(2r)^s \Gamma^s(1/p+1)}{|\det(\bar{\mathcal{S}})| \Gamma(s/p+1)} \prod_{i=s+1}^n (x_i^+ - x_i^-) \doteq V_{\mathcal{C}}. \quad (15)$$

We also remark that, by construction, for any  $r > 0$  and  $z \in \mathcal{H}(r)$

$$\Phi(z, r) = \mathcal{I}^{-1}(y) \cap \bar{\mathcal{C}}(z, r).$$

Note also that independent and identically distributed (iid) random samples inside  $\bar{\mathcal{C}}(z, r)$  can be easily obtained from iid uniform samples in the  $\ell_p$ -norm ball, whose generation is explained in [6]. Then, a probabilistic approximation of the volume of the intersection  $\Phi(z, r)$  may be computed by means of the randomized oracle presented in Algorithm 1, which is based on the uniform generation of iid samples in  $\bar{\mathcal{C}}(z, r)$ .

Then, note that the expected value of the random variable  $\hat{\phi}_N(z, r)$  with respect to the samples  $\chi^{(1)}, \dots, \chi^{(N)} \in \bar{\mathcal{C}}(z, r)$  is exactly the volume function  $\phi(z, r)$  appearing in (P-max-int)

$$\mathbb{E} \left[ \hat{\phi}_N(z, r) \right] = \phi(z, r).$$

---

#### Algorithm 1 Probabilistic Volume Oracle

---

##### 1. RANDOM GENERATION

Generate  $N$  iid uniform samples  $\zeta^{(1)}, \dots, \zeta^{(N)}$  in the  $s$ -dimensional ball  $\mathcal{B}(z, r)$

– For  $i = 1$  to  $N$

– Generate  $s$  iid scalars according to the unilateral Gamma density  $\gamma_j \sim G_{1/p, 1}$

– Construct the vector  $\eta \in \mathbb{R}^n$  of components  $\eta_j = s_j \gamma_j^{1/p}$ , where  $s_j$  are iid random signs

– Let  $\zeta^{(i)} = z + r w^{1/n} \frac{\eta}{\|\eta\|_p}$  where  $w$  is uniform in  $[0, 1]$

End for

Generate  $N$  iid uniform samples  $\xi^{(1)}, \dots, \xi^{(N)}$

– For  $i = 1$  to  $N$

– Generate  $\xi_j^i$  uniformly in the interval  $[x_{s+j}^-, x_{s+j}^+], j = 1, \dots, n-s$

End for

Construct the random samples in  $\bar{\mathcal{C}}(z, r)$  as follows

$$\chi^{(i)} = \begin{bmatrix} \bar{\mathcal{S}}^{-1} \zeta^{(i)} \\ \xi^{(i)} \end{bmatrix}, \quad i = 1, \dots, N$$

##### 2. CONSISTENCY TEST

– Compute the number of samples inside  $\mathcal{I}^{-1}(y)$  as follows

$$N_g = \sum_{i=1}^N \mathbb{I} \left( \|\mathcal{I} \chi^{(i)} - y\| \leq \rho \right)$$

##### 3. PROBABILISTIC ORACLE Return an approximation of the volume $\phi(z, r)$ as follows

$$\hat{\phi}_N(z, r) = \frac{N_g}{N} V_{\mathcal{C}}$$

where  $V_{\mathcal{C}}$  is defined in (15)

---

This immediately follows from the linearity of the expectation

$$\begin{aligned} \mathbb{E} \left[ \hat{\phi}_N(z, r) \right] &= \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \mathbb{I} \left( \chi^{(i)} \in \mathcal{I}^{-1}(y) \right) V_{\mathcal{C}} \right] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \mathbb{I} \left( \chi^{(i)} \in \mathcal{I}^{-1}(y) \right) \right] V_{\mathcal{C}}. \end{aligned}$$

Then, we have

$$\begin{aligned} \mathbb{E} \left[ \mathbb{I} \left( \chi^{(i)} \in \mathcal{I}^{-1}(y) \right) \right] &= 1 \cdot \text{Prob} \left\{ \chi^{(i)} \in \mathcal{I}^{-1}(y) \right\} + 0 \cdot \text{Prob} \left\{ \chi^{(i)} \notin \mathcal{I}^{-1}(y) \right\} \\ &= \text{vol}[\Phi(z, r)] / \text{vol}[\bar{\mathcal{C}}(z, r)] = \phi(z, r) / V_{\mathcal{C}}. \end{aligned}$$

Hence, we reformulate the problem (P-max-int) as the following stochastic optimization problem

$$\max_{z \in \mathcal{H}(r)} \mathbb{E} \left[ \hat{\phi}_N(z, r) \right].$$

This problem is classical and different stochastic approximation algorithms have been proposed, see for instance [18], [20] and references therein. In particular, in this paper, we use the SPSA (simultaneous perturbations stochastic approximation) algorithm, first proposed in [21], and further discussed in [23]. Convergence results under different conditions are detailed in the literature, see in particular the paper [16] which applies to non-differentiable functions. This approach is outlined in Algorithm 2.

---

**Algorithm 2** SPSA approach to (P-max-int)

---

1. INITIALIZATION Let  $k = 0$ , choose initial feasible center  $z_0 \in \mathcal{H}(r)$  and stepsize sequences  $\{a_k\}$ ,  $\{c_k\}$  satisfying  $a_k \rightarrow 0$ ,  $\sum_k a_k = \infty$ ,  $c_k \rightarrow 0$ ,  $\sum_k c_k = \infty$
  2. BERNOULLI GENERATION
    - Generate  $s$  iid Bernoulli points  $\Delta_k \in \{0, 1\}^s$
    - Define  $[\Delta_k^{-1}] \doteq [\Delta_{k,1}^{-1} \cdots \Delta_{k,n}^{-1}]^\top$
  3. APPROXIMATE GRADIENT
    - Compute the two perturbed values  $\hat{\phi}_N^\pm \doteq \hat{\phi}_N(z_k \pm c_k \Delta_k, r)$
    - Compute an approximate (sub)gradient as
$$\widehat{\partial}\hat{\phi}_N(z_k, r) = [\Delta_k^{-1}] \frac{\hat{\phi}_N^+ - \hat{\phi}_N^-}{2c_k}$$
  4. SUBGRADIENT ASCENT

$$z_{k+1} = z_k + a_k \widehat{\partial}\hat{\phi}_N(z_k, r)$$
  5. ITERATION Let  $k = k + 1$  and goto 2
- 

*Remark 1 (Scenario-based algorithms):* An alternative approach based on randomized methods can be also devised employing results on the scenario optimization method introduced in [5]. In particular, exploiting the results on discarded constraints, see [8], [10], an alternative algorithm can be constructed. The idea is as follows: (i) generate  $N$  samples  $\chi^{(i)}$  in  $\mathcal{I}^{-1}(y)$  according to the induced measure  $\tilde{\mu}_{\mathcal{I}^{-1}}$ , ii) solve the discarded-constraint random program

$$\min_{z, \gamma} \quad \gamma \quad (16)$$

$$\text{s.t.} \quad \frac{1}{L} \sum_{i \in I_L} \mathbb{I}(\|\mathcal{S}\chi^{(i)} - z\| \geq \gamma) \leq \epsilon \quad (17)$$

where  $I_L$  is a set of  $L$  indices constructed discarding in a prescribed way  $N - L$  indices from the set  $1, 2, \dots, N$ . Then, in [8], [10] it is shown how to choose  $N$  and the discarded set  $I_L$  to guarantee, with a prescribed level of confidence, that the result of optimization problem (16) is a good approximation of the true probabilistic radius  $r_0^{\text{pr}}(y, \epsilon)$ . However, this approach entails many technical difficulties, such as the random sample generation in point (i) and the optimal discarding procedure in point (ii), whose detailed analysis goes beyond the scope of this paper, and it is studied in [13]. We also point out that a different approach, also based on scenario optimization and discarded constraints, has been developed in [9] for identification and reliability problems, introducing the concept of interval predictor models.  $\diamond$

## V. A DETERMINISTIC RELAXATION TO (P-MAX-INT)

In this section, we propose a deterministic approach to (P-max-int) based on a semidefinite relaxation of the problem. We develop our approach focusing on the  $\ell_\infty$  norm; extensions to  $\ell_1$  and  $\ell_2$  norms are briefly discussed in Remark 2. First, note that in the case of  $\ell_\infty$  norms  $\mathcal{Q}$  is an hypercube of radius  $\rho$  and therefore  $\mathcal{I}^{-1}(y)$  is the polytope  $\mathcal{P}_X$  defined by the following linear inequalities

$$\begin{aligned} \mathcal{I}^{-1}(y) &= \{x \in \mathbb{R}^n \mid \|\mathcal{I}x - y\|_\infty \leq \rho\} \\ &= \left\{ x \in \mathbb{R}^n \mid \begin{bmatrix} \mathcal{I} \\ -\mathcal{I} \end{bmatrix} x \leq \begin{bmatrix} \rho \mathbf{1} + y \\ \rho \mathbf{1} - y \end{bmatrix} \right\} \end{aligned} \quad (18)$$

where  $\mathbf{1}$  is a vector of ones,  $\mathbf{1} = [1 \ 1 \ \dots \ 1]^\top$ . Since the exact computation of the volume of the intersection of two polytopic sets is in general costly and prohibitive in high dimensions, as discussed in Section III, we propose to maximize a suitably chosen lower bound of this volume. This lower bound can be computed as the solution of a convex optimization problem. The idea is to construct, for fixed  $r > 0$ , the maximal volume ellipsoid contained in the intersection  $\Phi(z, r)$ , which requires to solve the optimization problem

$$\begin{aligned} \max_{z, x_\mathcal{E}, P_\mathcal{E}} \quad & \text{vol}[\mathcal{E}(x_\mathcal{E}, P_\mathcal{E})] \\ \text{subject to} \quad & \mathcal{E}(x_\mathcal{E}, P_\mathcal{E}) \subseteq \Phi(z, r), \end{aligned} \quad (19)$$

where the ellipsoid of center  $x_\mathcal{E}$  and shape matrix  $P_\mathcal{E} \succeq 0$  is

$$\mathcal{E}(x_\mathcal{E}, P_\mathcal{E}) \doteq \{x \in \mathbb{R}^n \mid x = x_\mathcal{E} + P_\mathcal{E}w, \|w\|_2 \leq 1\}. \quad (20)$$

The problem of deriving the maximum volume ellipsoid inscribed in a polytope is a well-studied one, and concave reformulations based on linear matrix inequalities (LMI) are possible, see for instance [3], [2]. For completeness, we state a result in the next theorem.

*Theorem 1:* Let  $q \sim \mathcal{U}(\mathcal{Q})$  with  $\mathcal{Q} \equiv \mathcal{B}(\rho)$ , and  $\mathcal{S} = [\tilde{\mathcal{S}} \ 0_{s, n-s}]$ , with  $\tilde{\mathcal{S}} \in \mathbb{R}^{s, s}$ . Then, for given  $r > 0$ , a center that achieves a global optimum for problem (19) can be computed as the solution of the following semi-definite programming (SDP) problem

$$\begin{aligned} z_0^{\text{sdp}}(r) \in \arg_z \min_{z, x_\mathcal{E}, P_\mathcal{E}} \quad & -\log \det P_\mathcal{E} \\ \text{subject to} \quad & P_\mathcal{E} \succeq 0 \quad \text{and} \\ & \left[ \begin{array}{cc} (\rho + e_i^\top(y - \mathcal{I}x_\mathcal{E}))I_n & P_\mathcal{E}\mathcal{I}^\top e_i \\ \star & \rho + e_i^\top(y - \mathcal{I}x_\mathcal{E}) \end{array} \right] \succeq 0, \\ & i = 1, \dots, m \\ & \left[ \begin{array}{cc} (\rho - e_i^\top(y - \mathcal{I}x_\mathcal{E}))I_n & -P_\mathcal{E}\mathcal{I}^\top e_i \\ \star & \rho - e_i^\top(y - \mathcal{I}x_\mathcal{E}) \end{array} \right] \succeq 0, \\ & i = 1, \dots, m \\ & \left[ \begin{array}{cc} (r + \bar{e}_i^\top(z - \mathcal{S}x_\mathcal{E}))I_n & P_\mathcal{E}\mathcal{S}^\top \bar{e}_i \\ \star & r + \bar{e}_i^\top(z - \mathcal{S}x_\mathcal{E}) \end{array} \right] \succeq 0, \\ & i = 1, \dots, s \\ & \left[ \begin{array}{cc} (r - \bar{e}_i^\top(z - \mathcal{S}x_\mathcal{E}))I_n & P_\mathcal{E}\mathcal{S}^\top \bar{e}_i \\ \star & r - \bar{e}_i^\top(z - \mathcal{S}x_\mathcal{E}) \end{array} \right] \succeq 0, \\ & i = 1, \dots, s, \end{aligned}$$

where  $e_i$  and  $\bar{e}_i$  are elements of the canonical basis of  $\mathbb{R}^m$  and  $\mathbb{R}^s$ , respectively. Moreover, for all  $r > 0$ ,  $v_o^{\text{sdp}}(r) \geq v_o(r)$ , where we defined

$$v_o^{\text{sdp}}(r) \doteq 1 - \frac{\phi(z_o^{\text{sdp}}(r), r)}{\text{vol}[\mathcal{I}^{-1}(y)]}.$$

The proof of this result is available in the paper [12]. Note that, from this theorem, it follows that the SDP relaxation leads to a suboptimal violation function  $v_o^{\text{sdp}}(r)$ .

*Remark 2 (SDP relaxations for  $\ell_1$  and  $\ell_2$  norms):* An approach identical to that proposed in Theorem 1 can be developed for  $\ell_1$  norm, considering that also in this case the sets  $\mathcal{I}^{-1}(y)$  and  $\mathcal{C}(z, r)$  are a polytope and a cylinder with polytopic basis. Similarly, an analogous algorithm can be devised for (weighted)  $\ell_2$  norm. In this case, one should maximize the volume of an ellipsoid contained in the intersection of  $\mathcal{I}^{-1}(y)$  and  $\mathcal{C}(z, r)$ , which are respectively an ellipsoid and a cylinder with spherical basis, see [12] for a deeper analysis. Then, it can be easily seen, see e.g. [3], that this latter problem easily rewrites as a convex SDP optimization problem.  $\diamond$

## VI. NUMERICAL EXAMPLE

To study the performance of the (randomized and deterministic) algorithms previously presented, we consider the problem of estimating the parameters of a three-parameter finite-impulse response (FIR) model

$$y_k = x_1 u_k + x_2 u_{k-1} + x_3 u_{k-2} + q_k, \quad k = 1, \dots, m \quad (21)$$

where the input  $u_k$  is a known input sequence. The (unknown) nominal parameters were set to  $[1.25 \ 2.35 \ 0.5]^\top$ , and  $m = 200$  measurements were collected generating the input sequence  $\{u_k\}$  according to a Gaussian distribution with zero mean value and unit variance, and the measurement uncertainty  $q$  as a sequence of uniformly distributed noise with  $|q_k| \leq 0.5$ . Note that, in this example, we consider an identity solution operator  $\mathcal{S}$ , and thus  $X \equiv Y$  and the sets  $\mathcal{I}^{-1}(y)$  and  $\mathcal{S}\mathcal{I}^{-1}(y)$  coincide. That is, the solution coincides with the parameters, and the goal is to estimate  $z_i = x_i$ ,  $i = 1, 2, 3$ .

First, the optimal worst-case radius and the corresponding optimal solution have been computed by solving six linear programs (corresponding to finding the tightest box containing the polytope  $\mathcal{S}\mathcal{I}^{-1}(y)$ , see [19]). The computed worst-case optimal estimate is  $z_o^{\text{wc}} = [1.239 \ 1.788 \ 0.521]^\top$  and the worst-case radius is  $r_o^{\text{wc}}(y) = 0.0273$ , see Figure 1.

For comparison, we also computed the classical least-squares estimate  $\mathcal{A}_{\text{ls}}(y) = (\mathcal{I}^\top \mathcal{I})^{-1} \mathcal{I}^\top y = [1.265 \ 1.816 \ 0.542]^\top$ . Note that the least-square estimate is worst-case optimal as expected, and it is not even *interpolatory*, since it is outside the consistency set  $\mathcal{S}\mathcal{I}^{-1}(y)$  (see [25] for a formal definition of interpolatory algorithm).

Subsequently, in order to apply our probabilistic framework, we fix the accuracy level to  $\epsilon = 0.1$ , and aim at

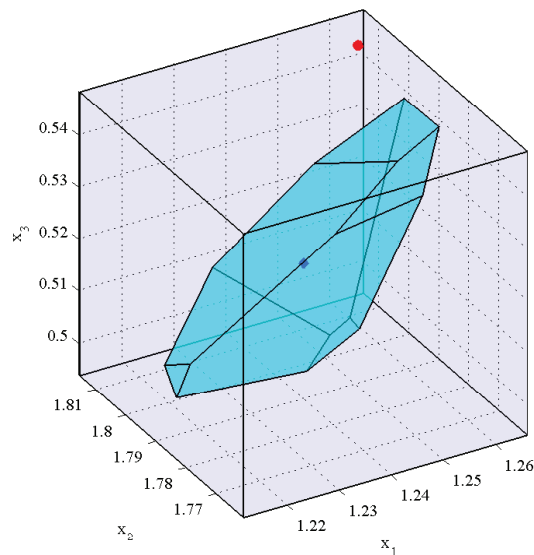


Fig. 1. Consistency set for the numerical example. The worst-case optimal radius  $r_o^{\text{wc}}(y)$  corresponds to the radius of the box enclosing the polytope  $\mathcal{S}\mathcal{I}^{-1}(y)$ . Its center, denoted by a diamond, represents the optimal worst-case estimate  $z_o^{\text{wc}}$ . The least-squares estimate is reported for completeness in the circle. Note that it falls outside of the polytope  $\mathcal{I}^{-1}(y)$ .

computing a probabilistic optimal radius and the corresponding optimal estimate according to definitions (4) and (5). Since the size is rather small ( $n = 3$ ), we use the techniques discussed in Section III for computing (P-max-int) exactly, obtaining the optimal violation function  $v_o(r)$  depicted in Figure 2.

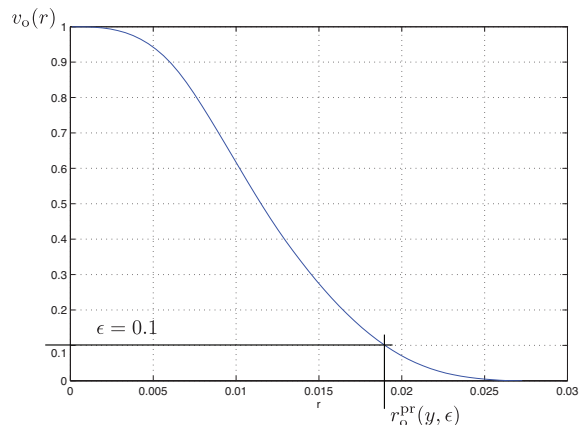


Fig. 2. Optimal violation function  $v_o(r)$ . To compute the probabilistic radius of information  $r_o^{\text{PF}}(y, \epsilon)$  corresponding to  $\epsilon = 0.1$ , we “invert” the plot as depicted in this figure.

By employing a simple bisection search algorithm over  $v_o(r)$ , the probabilistic radius of information was computed as  $r_o^{\text{PF}}(y, 0.1) = 0.0190$ . The corresponding optimal probabilistic estimate is given by  $z_o^{\text{PF}}(0.1) = [1.242 \ 1.790 \ 0.520]^\top$ . Note that the improvement in terms of radius of information is quite significant, being of the order of 30%.

The meaning of our approach is well explained in Figure

3. Indeed, in this figure we see that we look for the optimal “box” discarding a set of probability  $\epsilon = 0.1$ .

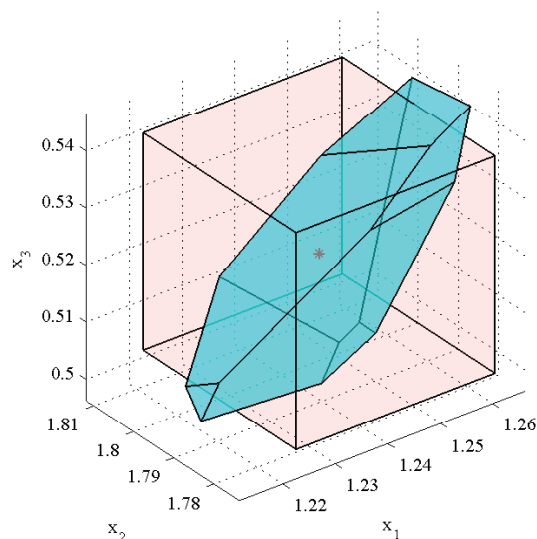


Fig. 3. Consistency set, and optimal “box” discarding a set of measure  $\epsilon = 0.1$ . The probabilistic optimal radius  $r_o^{\text{Pr}}(y, 0.1)$  corresponds to the radius of this box. The center, denoted by a star, represents the optimal probabilistic estimate  $z_o^{\text{Pr}}(y)$ .

Finally, to evaluate the quality of the randomized and deterministic relaxations introduced in Sections IV and V, we solved problem (P-max-int) for various values of radii in the interval  $[0.015, r_o^{\text{wc}}]$ , obtaining the results in Figure 4. It can be observed that the randomized approximation behaves quite well in this example, while the deterministic SDP performs poorly. Preliminary experiments seem to testify the fact that the quality of the SDP approximation increases as the dimensions grow, while the randomized algorithms tend to be less precise for larger dimensions, see e.g. the numerical example reported in [12].

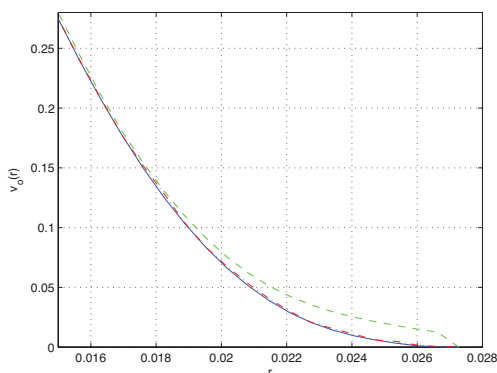


Fig. 4. Optimal violation function  $v_o(r)$  computed by the different algorithms presented in this paper: exact volume oracle (solid blue line), randomized algorithm (dashed red line) and SDP relaxation (dash-dotted green line).

## VII. CONCLUSIONS

In this paper, randomized and deterministic algorithms for computing the optimal violation function defined in [12] have

been presented. Their performance has been successfully tested on a numerical example of parameter estimation of a FIR(3) system.

## REFERENCES

- [1] H.K. Ahn, S-W. Cheng, and I. Reinbacher. Maximum overlap of convex polytopes under translation. *Algorithms and Computation*, 2010.
- [2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23:769–805, 1998.
- [3] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, 1994.
- [4] B. Bueler, A. Enge, and K. Fukuda. Exact volume computation for convex polytopes: a practical study. In G. Kalai and G. M. Ziegler, editors, *Polytopes Combinatorics and Computation*, volume 30, pages 131–154. Birkhauser, 2000.
- [5] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
- [6] G. Calafiore, F. Dabbene, and R. Tempo. Radial and uniform distributions in vector and matrix spaces for probabilistic robustness. In D.E. Miller and L. Qiu, editors, *Topics in Control and its Applications*, pages 17–31. Springer-Verlag, New York, 1999.
- [7] G. Calafiore, F. Dabbene, and R. Tempo. Research on probabilistic methods for control system design. *Automatica*, 47:1279–1293, 2011.
- [8] G.C. Calafiore. Random convex programs. *SIAM Journal on Optimization*, 20(6):3427–3464, 2010.
- [9] M.C. Campi, G.C. Calafiore, and S. Garatti. Interval predictor models: Identification and reliability. *Automatica*, 45(2):382–392, 1990.
- [10] M.C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2):257–280, 2011 (preliminary version available on Optimization Online, 2008).
- [11] F. Dabbene, M. Sznaier, and R. Tempo. A probabilistic approach to optimal estimation - Part I: Problem formulation and methodology. In *Proceedings IEEE Conference on Decision and Control*, 2012.
- [12] F. Dabbene, M. Sznaier, and R. Tempo. Probabilistic optimal estimation and filtering under uncertainty. *arXiv:1203.1429v2*, 2012.
- [13] F. Dabbene, M. Sznaier, and R. Tempo. A randomized algorithm for violation function approximation in system identification. *CNR - IEIIT Technical Report*, CNR-IEIIT-12-01-15-TO, 2012.
- [14] K. Fukuda and T. Uno. Polynomial-time algorithms for maximizing the intersection volume of polytopes. *Pacific Journal of Optimization*, 2007.
- [15] P.R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1950.
- [16] Y. He, M.C. Fu, and S. Marcus. Convergence of simultaneous perturbation stochastic approximation for nondifferentiable optimization. *IEEE Transactions on Automatic Control*, 48(8):1459–1463, 2003.
- [17] L.G. Khachiyan. Complexity of polytope volume computation. In J. Pach, editor, *New Trends in Discrete and Computational Geometry*, pages 91–101. Springer-Verlag, 1903.
- [18] H.J. Kushner and G.G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer-Verlag, New York, 2003.
- [19] M. Milanese and R. Tempo. Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control*, 30:730–738, 1985.
- [20] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 2003.
- [21] J.C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37:332–341, 1992.
- [22] J.C. Spall. Estimation via Markov Chain Monte Carlo. *IEEE Control Systems Magazine*, 23:34–45, 2003.
- [23] J.C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley, New York, 2003.
- [24] R. Tempo, G. Calafiore, and F. Dabbene. *Randomized Algorithms for Analysis and Control of Uncertain Systems*. Communications and Control Engineering Series. Springer-Verlag, London, 2005.
- [25] J.F. Traub, G.W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity*. Academic Press, New York, 1988.