A linear programming approach to online set membership parameter estimation for linear regression models

(Article begins on next page)

A linear programming approach to online set membership parameter estimation for linear regression models

M. Casini, A. Garulli, A. Vicino[1]

[1]Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche, Università di Siena, Via Roma 56, 53100 Siena, Italy, E-mail: {casini,garulli,vicino}@ing.unisi.it

Keywords: Set membership identification; recursive identification, set approximation, linear programming

Abstract

This paper presents a new technique for online set membership parameter estimation of linear regression models affected by unknown-but-bounded noise. An orthotopic approximation of the set of feasible parameters is updated at each time step. The proposed technique relies on the solution of a suitable linear program, whenever a new measurement leads to a reduction of the approximating orthotope. The key idea for preventing the size of the linear programs from steadily increasing is to propagate only the binding constraints of these optimization

problems. Numerical studies show that the new approach outperforms existing recursive set approximation techniques, while keeping the required computational burden within the same order of magnitude.

1 Introduction

Recursive parameter estimation is a crucial issue in many different contexts, including system identification, adaptive control and fault diagnosis [1, 2, 3]. The classical Recursive Least Squares (RLS) solution and its numerous variations have been studied and widely employed in a *stochastic* framework, in which the noise affecting the data is modeled as a stochastic process. As a consequence, the parameter estimates are themselves random variables and the uncertainty associated to the estimates can be described in probabilistic terms via the corresponding probability density function. An alternative approach, often referred to as *deterministic*, assumes that noise signals are Unknown-But-Bounded (UBB) in some norm. This naturally leads to the formulation of the so-called *set membership* estimation framework, in which the the uncertainty associated to the parameters to be estimated is described by the *Feasible Parameter Set (FPS)*. This is the set of all parameter vectors that are compatible with the available data and the UBB noise assumption.

The main idea of the set membership approach dates back to the early works of Bertsekas and Rhodes [4], and Schweppe [5]. Starting from the 80s, a number of different approaches and techniques have been proposed to tackle a wide variety of estimation problems in this framework (see e.g., [6, 7, 8, 9, 10, 11, 12, 13] and references therein). A great amount of work has been devoted to find tractable approximations of the FPS, taking into account that the shape and complexity of such a set

depends both on the norm used to bound the noise and on the structure of the selected model class. In the literature, the most popular setting is by far that of linear regression models with ℓ_∞ bounded noise, in which the FPS turns out to be a convex polytope. On the other hand, nonconvex FPSs usually arise whenever one deals with switching systems [14, 15, 16], nonlinear models [17, 18, 19], quantized measurements [20, 21], or LPV systems [22].

Since the true FPS is usually too complex to be characterized exactly, recursive approximations of the FPS through simply shaped regions have been intensively investigated, along with the computation of nominal estimates satisfying some properties related to the FPS (e.g., the Chebishev center of the approximating set, in some norm). Several classes of approximating regions have been considered in the literature, including orthotopes [23, 24, 25], ellipsoids [26, 27, 28], parallelotopes [29, 30, 31], zonotopes [32, 33], orthonormal basis functions [34] and others. While recursive algorithms employing orthotopes often provide very coarse approximations of the true feasible set, parallelotopes have proven to return tighter estimates with respect to ellipsoids, especially when the number of available data is strongly limited (see the comparative study in [35]). On the other hand, the computational complexity of these techniques turns out to be of the same order of the RLS algorithm. In [24], a recursive orthotopic approximation is computed by keeping track of a suitable subset of the constraints defining the true FPS. Nevertheless, all the recursive approximations are generally much more conservative than the batch orthotopic approximation, which provides the exact uncertainty interval for each parameter and can be obtained by solving $2n$ Linear Programs (LPs), where n is the number of parameters to be estimated. Clearly, this approach is not feasible in an online estimation framework, because the number of constraints in the LPs grows over time.

Since the seminal work of Karmarkar [36], which introduced an algorithm for solving LPs in

polynomial time, it became quickly clear that one could exploit the steadily increasing computational power of modern computers to solve problems of very large dimension. Nowadays, a variety of efficient solvers is available and it can be stated that solving LPs is "a mature technology" [37], just like it is commonly believed for least squares problems. The main idea at the basis of the present work is to exploit this technology to improve the online approximation of the FPS, with respect to available recursive approaches, while keeping the required computational burden in the same order of magnitude as that of the RLS algorithm. In particular, a new technique providing an orthotopic approximation of the FPS is presented. The approach relies on a constraint selection technique which builds on the concept of *binding constraints* in LPs. The approach relies on a constraint selection technique, which propagates a data structure similar to that of [24], but builds on the more powerful concept of binding constraints in LPs. A preliminary version of these results has been presented in [38].

The paper is organized as follows. A motivating example is first introduced in Section 2, to illustrate the trade-off between the quality of the FPS approximations and the required computational burden. Preliminaries on LPs are briefly recalled in Section 3, while the main problem is formulated in Section 4. The new approximation technique is presented in Section 5. Section 6 reports numerical tests which compare the proposed approach with existing recursive estimation algorithms. Finally, concluding remarks are given in Section 7.

2 Motivating example

Let us recall the following example that has been used as a benchmark in several papers addressing the problem of set membership parameter estimation (see, e.g., [26, 29]). Consider the ARX model

$$y(t) + 1.3y(t-1) + 0.4y(t-2) = u(t) + 0.8u(t-1) + e(t) \tag{1}$$

where the UBB noise $e(t)$ satisfies $|e(t)| \leq \delta$. By casting the model in the regression form

$$y(t) = \varphi^T(t)\theta + e(t) \tag{2}$$

where $\varphi(t) = [-y(t-1)\ -y(t-2)\ u(t)\ u(t-1)]$ and $\theta \in \mathbb{R}^4$ is the vector of parameters to be estimated, one has that the *Feasible Parameter Set (FPS)*, i.e. the set of all parameter vectors compatible with the model equation and the UBB noise assumption, is defined by the inequalities

$$-\delta \leq y(t) - \varphi^T(t)\theta \leq \delta\ , \quad t = 1, \ldots, N \tag{3}$$

where N is the number of available input and output measurements. Clearly, the inequalities (3) define a polytope in the space of parameter vectors θ. Thus, if one is interested in estimating the feasible interval for each parameter θ_i, $i = 1, \ldots 4$, the minimum box containing such a polytope has to be computed. This requires to solve 8 linear programming (LP) problems, each one with $2N$ constraints.

The computation of the minimum outer box is clearly infeasible in online estimation problems, because the number of constraints grows linearly over time. For this reason, approximations of the FPS based on simple regions have become common practice in recursive set membership estimation. Unfortunately, it turns out that these approximations are usually much coarser that the minimum

outer box. Figure 1 shows that after $N = 500$ iterations, the volume of the minimum box turns out to be two orders of magnitude smaller than the volume of the recursive ellipsoidal and parallelotopic approximations, proposed respectively in [26] and [29]. On the other hand, the number of constraints bounding the FPS is depicted in Figure 2. It can be noticed that only few constraints are important for the computation of the minimum outer box, but keeping track of all of them is computationally demanding. Such results have been obtained by averaging over 100 different runs of the considered algorithms, with white input signals $u(t)$ and $e(t)$, uniformly distributed in the intervals $[-1, 1]$ and $[-0.1, 0.1]$, respectively. Anyway, very similar results are obtained with different input and noise choices. This suggests that it would be desirable to devise an approximation technique, able to provide an online approximation of the FPS as close as possible to the minimum outer box, but requiring a computational burden comparable to that of the classical recursive approaches based on ellipsoids or parallelotopes. This is precisely the objective of the present paper.

3 Preliminaries on linear programs

In this section, some basic properties of linear programs that will be exploited in the paper are briefly recalled. Consider the LP problem

$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & \\ Ax \leq b & \end{array} . \tag{4}$$

Let a_i^T denote the ith row of matrix A. The following definitions are introduced:

i) $\mathcal{X} = \{x |\ Ax \leq b\}$ is the *constraint set* of the LP (4).

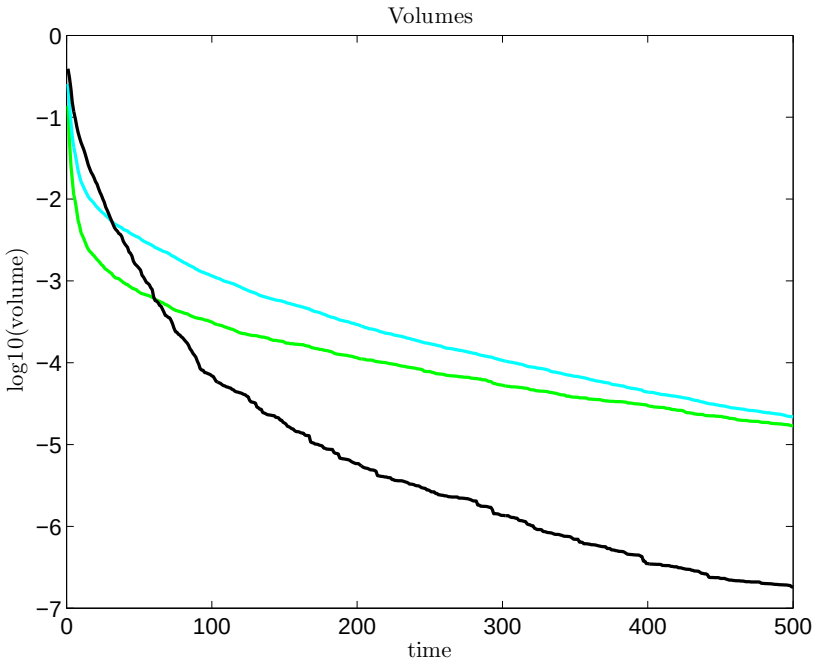


Figure 1: Average log-volumes of regions containing the FPS: ellipsoidal approximation (cyan), parallelotopic approximation (green), minimum outer box (black).

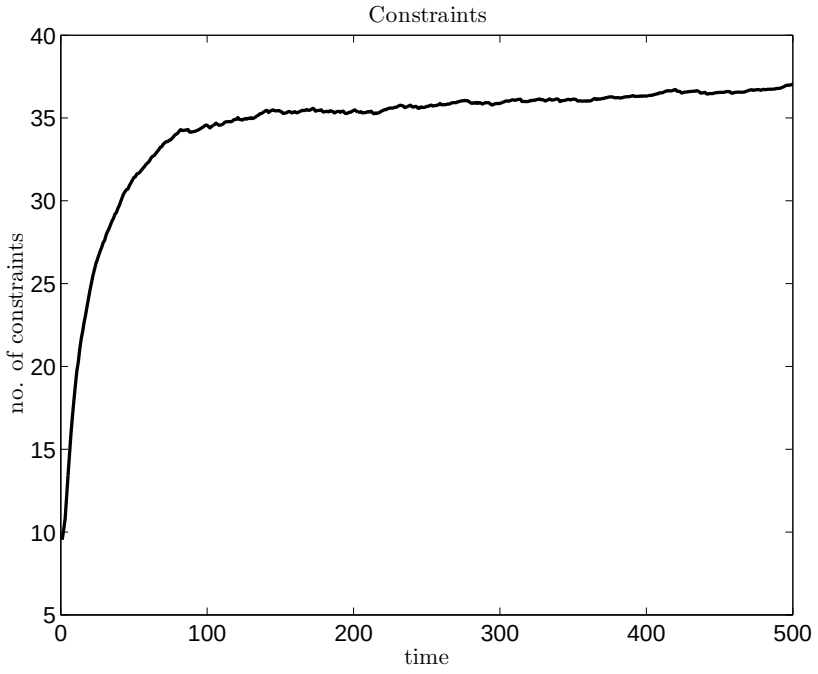


Figure 2: Average number of constraints bounding the FPS.

ii) $\Xi = \{x \in \mathcal{X} |\ x = \arg\max c^T x\}$ is the *solution set* of the LP (4).

iii) Let $\bar{x} \in \mathcal{X}$; the constraint $a_i^T x \leq b_i$ is an *active constraint* at $\bar{x}$, if $a_i^T \bar{x} = b_i$.

iv) The constraint $a_i^T x \leq b_i$ is a *binding constraint* of the LP (4) if there exists $x^* \in \Xi$ such that $a_i^T x^* = b_i$.

v) Let $\mathcal{I}$ be the set of indexes i such that $a_i^T x \leq b_i$ is a binding constraint. The set

$$\mathcal{A} : \{x |\ a_i^T x \leq b_i,\ i \in \mathcal{I}\}$$

is the *binding set* of the LP. Clearly, $\mathcal{X} \subseteq \mathcal{A}$.

The following properties follow directly from the above definitions.

Proposition 1. *Consider the LP*

$$\begin{array}{ll} \max & c^T x \\ s.t. & \\ x \in \mathcal{A} & \end{array} \tag{5}$$

where $\mathcal{A}$ is the binding set of the LP (4). Then, the LPs (4) and (5) have the same solution and the same solution set Ξ.

Proposition 2. *Let $\mathcal{H} = \{x |\ \overline{a}^T x \leq \overline{b}\}$. Then,*

$$\begin{array}{lll} \max & c^T x & \geq \\ s.t. & & \\ x \in \mathcal{A} \bigcap \mathcal{H} & & \end{array} \quad \begin{array}{ll} \max & c^T x \\ s.t. & \\ x \in \mathcal{X} \bigcap \mathcal{H} & \end{array} . \tag{6}$$

Loosely speaking, Proposition 1 states that replacing the constraint set with the binding set does not introduce any conservatism in the solution of a single LP. However, Proposition 2 highlights that the intersection with a new constraint $\mathcal{H}$ may lead to a conservative solution if only the binding constraint of the original LP are propagated.

4 Problem formulation

In this section, the problem of computing an online approximation of the FPS is formally stated. Consider the linear regression model

$$y(t) = \varphi^T(t)\theta + e(t) \tag{7}$$

where $\theta \in \mathbb{R}^n$ is the vector of parameters to be estimated, $\varphi(t) \in \mathbb{R}^n$ is a known time-varying regression vector, and $e(t)$ is an unknown-but-bounded noise signal such that

$$|e(t)| \leq \delta, \quad \forall t. \tag{8}$$

Then, the feasible parameter set at a generic time t is defined as

$$\Theta(t) = \{\theta \in \Theta_0 : |y(k) - \varphi^T(k)\theta| \leq \delta, \quad k = 1, \ldots, t\} = \bigcap_{k=1}^{t} \mathcal{S}(k) \tag{9}$$

where $\mathcal{S}(k) = \{\theta : |y(k) - \varphi^T(k)\theta| \leq \delta\}$ is a "strip" in the parameter space and Θ_0 represents the a priori knowledge on the parameters to be estimated. The set $\Theta(t)$ is a polytope in the parameter space. The aim of the paper is to solve the following problem.

Problem 1. *At each time t, compute a set $\mathcal{R}(t)$ such that $\mathcal{R}(t) \supseteq \Theta(t)$.*

Clearly, the solution set $\mathcal{R}(t)$ should satisfy the following conflicting requirements: i) it must be as "small" as possible according to some measure of size (e.g., the volume); ii) the computational burden required to compute $\mathcal{R}(t)$ should not grow over time.

A standard solution of Problem 1 is based on axis-aligned boxes, also known as *orthotopes*. This class of regions is defined as

$$\mathcal{O}(\overline{\theta}, d) = \{\theta : \ \theta = \overline{\theta} + \mathrm{diag}(d)w, \ \|w\|_\infty \leq 1\}.$$

where $\overline{\theta}, d, w \in \mathbb{R}^n$, $d_i \geq 0$, $i = 1, \ldots, n$, and $\mathrm{diag}(d)$ is a diagonal matrix with diagonal equal to d. The sets

$$\mathcal{F}_i = \{\theta \in \mathcal{O}(\overline{\theta}, d) : \ \theta_i = \overline{\theta}_i + d_i\} \ , \quad \mathcal{F}_{i+n} = \{\theta \in \mathcal{O}(\overline{\theta}, d) : \ \theta_i = \overline{\theta}_i - d_i\} \ , \quad i = 1, \ldots, n,$$

are the $(n-1)$-dimensional *faces* of the orthotope. Let the vectors e_i, $i = 1, \ldots, n$ denote the columns of the identity matrix. Then, the minimum volume orthotope containing the polytope $\Theta(t)$, hereafter denoted by $\mathcal{O}^*(\Theta(t))$, can be computed by solving the $2n$ LPs

$$\begin{array}{rcl} \beta_i(t) & = & \max \quad e_i^T\theta \ , \\ & & \text{s.t.} \\ & & \theta \in \Theta(t) \end{array} \qquad \begin{array}{rcl} \beta_{i+n}(t) & = & \min \quad e_i^T\theta \\ & & \text{s.t.} \\ & & \theta \in \Theta(t) \end{array} \tag{10}$$

for $i = 1, \ldots, n$. Then, one has that $\mathcal{O}^*(\Theta(t)) = \mathcal{O}(\overline{\theta}^*(t), d^*(t))$, where

$$\overline{\theta}_i^*(t) = \frac{\beta_i(t) + \beta_{i+n}(t)}{2} \ , \qquad d_i^*(t) = \frac{\beta_i(t) - \beta_{i+n}(t)}{2} \ , \quad i = 1, \ldots, n.$$

While it is apparent that $\mathcal{O}^*(\Theta(t))$ is a tight approximation of $\Theta(t)$, and that its faces provide the exact feasible intervals for each parameter θ_i, it is also clear that from the computational viewpoint

this is not a feasible approach in online estimation problems, because the number of constraints in the constraint sets $\Theta(t)$ of the LPs (10) is equal to $2t$, and therefore it grows linearly over time.

In the literature, a number of recursive approximation techniques of the FPS have been proposed, based on the following scheme

$$\mathcal{R}(t+1) \supseteq \mathcal{R}(t) \bigcap \mathcal{S}(t+1) \tag{11}$$

where $\mathcal{R}(t)$ belongs to a specified class of approximating regions $\mathcal{R}$. If the scheme is initialized with a set $\mathcal{R}(0) \supseteq \Theta_0$, then from (9) and (11) one has that $\mathcal{R}(t) \supseteq \Theta(t)$, $\forall t$, i.e., $\mathcal{R}(t)$ is a solution of Problem 1. The techniques proposed in the literature employ several different classes of approximating regions $\mathcal{R}$, such as orthotopes, ellipsoids, parallelotopes, zonotopes, and others. The criteria adopted to compute $\mathcal{R}(t+1)$ in (11) are usually based on the minimization of a suitable measure of the size of $\mathcal{R}(t+1)$, like the volume or the diameter. Nevertheless, both the choice of the set class $\mathcal{R}$ and of the size measure are most often dictated by the need of keeping the computational burden limited and approximately constant at each time t, to the detriment of the quality of the approximation, as shown in the example presented in Section 2. The approach proposed in the next section aims at providing a much better trade-off between computational complexity and quality of the FPS approximation.

5 Orthotopic approximation technique

Let us consider the feasible parameter set $\Theta(t)$, defined according to (9), and the corresponding minimum bounding orthotope $\mathcal{O}^*(\Theta(t))$, obtained by computing the solutions $\beta_i(t)$ of the $2n$ LPs (10). The next result provides the theoretical basis for a strategy aimed at selecting the most significant constraints in the LPs.

Proposition 3. *Let $\mathcal{A}_i(t)$, $i = 1, \ldots, 2n$, be the binding sets of the $2n$ LPs* (10) *and define*

$$\mathcal{A}(t) = \bigcap_{i=1}^{2n} \mathcal{A}_i(t). \tag{12}$$

Let $\Xi_i(t)$ $i = 1, \ldots, 2n$, be the solution sets of the $2n$ LPs (10) *and $v^{(i)}(t) \in \Xi_i(t)$ be an element of the ith solution set. Then, the following statements hold.*

i) $\mathcal{O}^*(\Theta(t)) = \mathcal{O}^*(\mathcal{A}(t))$.

ii) $\mathcal{O}^*(\Theta(t) \bigcap \mathcal{S}(t+1)) \subseteq \mathcal{O}^*(\mathcal{A}(t) \bigcap \mathcal{S}(t+1))$.

iii) If $v^{(i)}(t) \in \mathcal{S}(t+1)$ for some $1 \leq i \leq 2n$, then

$$\beta_i(t+1) = \beta_i(t).$$

iv) If $v^{(i)}(t) \in \mathcal{S}(t+1)$, for all $i = 1, \ldots, 2n$, then

$$\mathcal{O}^*(\Theta(t+1)) = \mathcal{O}^*(\Theta(t)).$$

v) If $\Xi_i(t) = \{v^{(i)}(t)\}$ and $v^{(i)}(t) \notin \mathcal{S}(t+1)$, then $\beta_i(t+1) < \beta_i(t)$ if $1 \leq i \leq n$, and $\beta_i(t+1) > \beta_i(t)$ if $n+1 \leq i \leq 2n$.

Proof. Item i) and ii) follow from Proposition 1 and 2, respectively, applied to the $2n$ LPs (10).

iii) Let us first observe that $v^{(i)}(t) \in \Xi_i(t)$ implies $v_i^{(i)}(t) = \beta_i(t)$. Being $v^{(i)}(t) \in \mathcal{S}(t+1)$, one has that $v^{(i)}(t) \in \Theta(t) \bigcap \mathcal{S}(t+1) = \Theta(t+1)$ and hence there exists an element of $\Theta(t+1)$ whose ith coordinate is equal to $\beta_i(t)$. Hence, one necessarily has that $\beta_i(t+1) = \beta_i(t)$.

iv) It is a straightforward consequence of item iii).

v) Let i such that $1 \leq i \leq n$ (the case $n+1 \leq i \leq 2n$ is analogous). Being $v^{(i)}(t)$ the unique element in the solution set of the leftmost LP in (10), one has

$$\beta_i(t) = e_i^T v^{(i)}(t) > e_i^T z \ , \quad \forall z \in \Theta(t), \ z \neq v^{(i)}(t).$$

Since $v^{(i)}(t) \notin \mathcal{S}(t+1)$, then also $v^{(i)}(t) \notin \Theta(t+1)$ and therefore

$$\begin{array}{ccccc} \beta_i(t) & > & \max & e_i^T z & = \ \beta_i(t+1) \\ & & \text{s.t.} & & \\ & & z \in \Theta(t+1) & & \end{array}$$

which concludes the proof. □

It is worth observing that in general the solution set $\Xi_i(t)$ of the ith LP can contain infinite elements. This occurs whenever the corresponding active set contains a constraint orthogonal to the vector e_i. The most common situation is however the one in which $\Xi_i(t)$ is a singleton, i.e. $\Xi_i(t) = \{v^{(i)}(t)\}$. In such a case, items iii) and v) in Proposition 3 lead to the conclusions that the minimum outer orthotope containing $\Theta(t)$ is reduced if and only if $v^{(i)}(t) \notin \mathcal{S}(t+1)$. Therefore, the key idea underlying the approach proposed hereafter is to use the elements $\{v^{(i)}(t)\}$ as *markers* to detect whether the FPS approximation has to be updated or not.

The results in Proposition 3 can be exploited to devise a new technique for computing an online approximation of the FPS. The objective would be to propagate over time only the constraints of the feasible parameter set $\Theta(t)$ which belong also to the binding set $\mathcal{A}(t)$, and then compute the approximating orthotope only with respect to these constraints. Due to item i) in Proposition 3, this would lead to compute the minimum orthotope $\mathcal{O}^*(\Theta(t))$, but would still require to solve $2n$ LPs at each time step. Hence, one can exploit the markers $\{v^{(i)}(t)\}$ to decide whether a face of the

approximating orthotope must be updated or not, according to items iii)-v) in Proposition 3. Since the proposed method propagates only the binding constraints, the computed orthotope turns out to be an outer approximation of the optimal one, due to item ii).

In order to illustrate the idea, let us assume that at a generic time t the set $\mathcal{A}(t)$ defined in (12) is available, along with $2n$ elements $v^{(i)}(t)$, $i = 1, \ldots, 2n$, such that $v^{(i)}(t)$ belongs to the solution set of the LP

$$\begin{array}{ll} \max\,(\min) & e_i^T\theta \\ \text{s.t.} & \\ \theta \in \mathcal{A}(t) & \end{array} \quad . \tag{13}$$

where the shorthand notation max(min) will be adopted henceforth, meaning that max applies for $i = 1, \ldots, n$ and min for $i = n+1, \ldots, 2n$. When a new measurement is processed at time $t+1$, one has that the minimum orthotope containing the polytope defined by the intersection between the set $\mathcal{A}(t)$ and the new "strip" $\mathcal{S}(t+1)$ is in general larger than the minimum orthotope containing the true FPS, according to item ii) in Proposition 3. Nevertheless, item iii) in the same Proposition states that if an element of the solution set of the ith LP in (13) belongs to $\mathcal{S}(t+1)$, the new measurement does not modify the solution of the ith LP. Therefore, only the LPs corresponding to indices i such that $v^{(i)}(t) \notin \mathcal{S}(t+1)$ have to be solved at each time step. In particular, if all elements $v^{(i)}(t) \in \mathcal{S}(t+1)$, $i = 1, \ldots, 2n$, there is no need to update the bounding orthotope.

According to the above observations, the following procedure is proposed for updating an orthotope containing the feasible set $\Theta(t)$ at each time t.

Procedure 1. *Step 1 [Data structure]. At a generic time t, let:*

i) $\mathcal{C}(t)$ be a polytope whose constraints are a subset of the constraints of $\Theta(t)$ (hence, one has

$\mathcal{C}(t) \supseteq \Theta(t)$);

ii) $\mathcal{O}(t) = \mathcal{O}(\overline{\theta}(t), d(t)) = \mathcal{O}^*(\mathcal{C}(t))$ *be the minimum orthotope containing* $\mathcal{C}(t)$*;*

iii) $v^{(i)}(t)$, $i = 1, \ldots, 2n$, *be* $2n$ *vectors such that* $v^{(i)}(t) \in \mathcal{C}(t) \bigcap \mathcal{F}_i(t)$, *where* $\mathcal{F}_i(t)$ *is the* i*th face of* $\mathcal{O}(t)$. *In other words,* $v^{(i)}(t)$ *is an element of* $\mathcal{C}(t)$ *belonging to the* i*th face of* $\mathcal{O}(t)$, *i.e. satisfying*

$$v_i^{(i)}(t) = \overline{\theta}_i(t) + d_i(t) \ , \quad v_i^{(i+n)}(t) = \overline{\theta}_i(t) - d_i(t)$$

for $i = 1, \ldots, n$.

Let us denote by $\mathcal{C}_i(t)$ *the set defined by those constraints in* $\mathcal{C}(t)$ *that are active at least at one element of* $\mathcal{C}(t) \bigcap \mathcal{F}_i(t)$.

Step 2 [LPs solution]. For each $i = 1, \ldots, 2n$, *if* $v^{(i)}(t) \notin \mathcal{S}(t+1)$ *set*

$$\begin{aligned} v^{(i)}(t+1) \quad = \quad & \arg\max\,(\min) \quad e_i^T \theta \\ & s.t. \\ & \theta \in \mathcal{C}(t) \bigcap \mathcal{S}(t+1) \end{aligned} \tag{14}$$

and $\mathcal{C}_i(t+1) = \mathcal{A}_i(t+1)$, *where* $\mathcal{A}_i(t+1)$ *is the binding set of the LP* (14).

Otherwise, if $v^{(i)}(t) \in \mathcal{S}(t+1)$, *set* $v^{(i)}(t+1) = v^{(i)}(t)$ *and* $\mathcal{C}_i(t+1) = \mathcal{C}_i(t)$. *Then, let*

$$\mathcal{C}(t+1) = \bigcap_{i=1}^{2n} \mathcal{C}_i(t+1). \tag{15}$$

Step 3 [Orthotope update]. Compute the new approximating orthotope

$$\mathcal{O}(t+1) = \mathcal{O}(\overline{\theta}(t+1), d(t+1)),$$

where

$$\overline{\theta}_i(t+1) = \frac{v_i^{(i)}(t+1) + v_i^{(i+n)}(t+1)}{2} \tag{16}$$

$$d_i(t+1) = \frac{v_i^{(i)}(t+1) - v_i^{(i+n)}(t+1)}{2} \tag{17}$$

for $i = 1, \ldots, n$. Set $t = t+1$, and go back to Step 1.

The following result proves that Procedure 1 can be indeed iterated and it returns a solution to Problem 1, in terms of a sequence of nested orthotopes.

Proposition 4. *Let the a priori set Θ_0 be a closed polytope, $\mathcal{C}(0)$ be any closed polytope whose constraints are a subset of the constraints of Θ_0, and $\mathcal{O}(0) = \mathcal{O}^*(\mathcal{C}(0))$. Let $v^{(i)}(0) \in \mathcal{C}(0) \bigcap \mathcal{F}_i(0)$, $i = 1, \ldots, 2n$, where $\mathcal{F}_i(0)$ is the ith face of $\mathcal{O}(0)$. Then, by applying Procedure 1 at any time t, one has:*

i) The sets $\mathcal{C}(t+1)$, $\mathcal{O}(t+1)$ and the vectors $v_i^{(i)}(t+1)$ satisfy the same properties as $\mathcal{C}(t)$, $\mathcal{O}(t)$ and $v_i^{(i)}(t)$, stated in Step 1 of Procedure 1.

ii) $\mathcal{O}(t+1) \subseteq \mathcal{O}(t)$, $\forall t$.

iii) $\mathcal{O}(t) \supseteq \Theta(t)$, $\forall t$.

Proof. i) Since $\mathcal{C}(t) \supseteq \Theta(t)$ and $\Theta(t+1) = \Theta(t) \bigcap \mathcal{S}(t+1)$, one has $\mathcal{C}(t) \bigcap \mathcal{S}(t+1) \supseteq \Theta(t+1)$. Moreover, the constraints of the binding sets $\mathcal{A}_i(t+1)$ of the LPs (14) are a subset of those of $\mathcal{C}(t) \bigcap \mathcal{S}(t+1)$, and hence of $\Theta(t+1)$. Being $\mathcal{C}_i(t+1)$ equal either to $\mathcal{A}_i(t+1)$ or to $\mathcal{C}_i(t)$, from (15) one can conclude that the constraints of $\mathcal{C}(t+1)$ are a subset of those of $\Theta(t+1)$ and therefore $\mathcal{C}(t+1) \supseteq \Theta(t+1)$.

Let us now consider the vectors $v^{(i)}(t+1)$, for $i = 1, \ldots, n$ (the case $i = n+1, \ldots, 2n$ is analogous). For each index i such that $v^{(i)}(t) \notin \mathcal{S}(t+1)$, one has that $v^{(i)}(t+1)$ belongs to the binding sets $\mathcal{A}_j(t+1)$ of the LPs (14), for all $j = 1, \ldots, 2n$. Therefore, $v^{(i)}(t+1) \in \mathcal{C}(t+1)$. From (16)-(17), one gets $v_i^{(i)}(t+1) = \overline{\theta}_i(t+1) + d_i(t+1)$, i.e. $v_i^{(i)}(t+1) \in \mathcal{F}_i(t+1)$. Hence, $v^{(i)}(t+1) \in \mathcal{C}(t+1) \bigcap \mathcal{F}_i(t+1)$. Moreover, being $\mathcal{C}(t+1) \subseteq \mathcal{A}_i(t+1)$ one has

$$\begin{array}{lcl} \max & e_i^T\theta \leq & \max \quad e_i^T\theta = v_i^{(i)}(t+1) = \overline{\theta}_i(t+1) + d_i(t+1) \\ \text{s.t.} & & \text{s.t.} \\ \theta \in \mathcal{C}(t+1) & & \theta \in \mathcal{A}_i(t+1) \end{array}$$

On the other hand, since $v^{(i)}(t+1) \in \mathcal{C}(t+1)$,

$$\begin{array}{l} \max \quad e_i^T\theta \geq v_i^{(i)}(t+1) = \overline{\theta}_i(t+1) + d_i(t+1) \\ \text{s.t.} \\ \theta \in \mathcal{C}(t+1) \end{array}$$

By repeating the reasoning for every dimension i, one can conclude that $\mathcal{O}(t+1) = \mathcal{O}^*(\mathcal{C}(t+1))$.

ii) According to Steps 2 and 3 of Proposition 1, the ith face of $\mathcal{O}(t+1)$ is updated only if $v^{(i)}(t) \notin \mathcal{S}(t+1)$. In such a case, for $i = 1, \ldots, n$ one has

$$\begin{array}{lcl} v_i^{(i)}(t+1) = & \max \quad e_i^T\theta \leq & \max \quad e_i^T\theta = v_i^{(i)}(t) \\ & \text{s.t.} & \text{s.t.} \\ & \theta \in \mathcal{C}(t) \bigcap \mathcal{S}(t+1) & \theta \in \mathcal{C}(t) \end{array}$$

A similar reasoning applies to the minimum optimization problems for $i = n+1, \ldots, 2n$, and hence one gets $\mathcal{O}(t+1) \subseteq \mathcal{O}(t)$.

iii) It is a straightforward consequence of the fact that at each time t, $\mathcal{O}(t) = \mathcal{O}^*(\mathcal{C}(t))$ and $\mathcal{C}(t) \supseteq \Theta(t)$. $\square$

In order to initialize Procedure 1, a possible choice is to use the first t_0 measurements to compute the minimum orthotope containing the true feasible set $\Theta(t_0)$ at a certain time t_0, by solving the corresponding $2n$ LPs. Then, one can select $v^{(i)}(t_0)$, $i = 1, \ldots, 2n$, by picking any element from the solution set of the ith LPs and set $\mathcal{C}(t_0) = \mathcal{A}(t_0)$ defined by (12).

The rationale behind Procedure 1 is that ideally one would like to propagate the binding set $\mathcal{A}(t)$, because it contains those constraints that are sufficient to yield the minimum bounding orthotope $\mathcal{O}^*(\Theta(t))$. Hence, each time one face of the approximating orthotope is tightened, the corresponding LP is solved and the binding set of that LP is inserted in the constraint set $\mathcal{C}(t)$. Unfortunately, the recursive updating of the constraint set does not guarantee that $\mathcal{C}(t)$ coincides with $\mathcal{A}(t)$ at a generic time t, because a constraint that is not binding at a certain time, might become binding later due to intersection with a new constraint. Nevertheless, the numerical tests reported in the next section testify that the procedure is able to provide a much tighter approximation with respect to classical recursive methods, while the number of LPs to be solved and of constraints to be kept track of, turns out to be much smaller with respect to those required by the computation of the minimum orthotope $\mathcal{O}^*(\Theta(t))$.

It is worthwhile to remark that besides the outer bounding orthotope, the proposed procedure provides also a polytopic approximation $\mathcal{C}(t)$ of the true feasible set. In this respect, the proposed method can be considered among set membership techniques based on limited-complexity polytopic approximations (see, e.g., [39, 24, 40] to cite a few).

6 Numerical tests

In this section, Procedure 1 is compared to three recursive approaches for the approximation of the FPS: the Recursive Optimal Bounding Ellipsoidal (ROBE) algorithm, in the version firstly proposed in [26], the Recursive Optimal Bounding Parallelotopic (ROBP) algorithm, introduced in [29], and the Recursive Central Estimate (RCE), presented in [24]. The ROBE and ROBP algorithms follow the recursive scheme (11) and at each time step minimize the volume of the approximating ellipsoid and parallelotope, respectively. The RCE adopts an orthotopic set representation similar to the one used in this paper. However, to update the orthotope, it propagates $2n$ subsets of inequalities, one for each side of the approximating orthotope, each one containing n constraints. On the contrary, the technique proposed in this paper relies on a unique set of constraints, containing all the binding constraints of the solved LPs. Moreover, for each update of the approximating orthotope, the proposed technique requires the solution of a single LP, while the RCE algorithm requires the computation of the inverse of n matrices of dimension $n \times n$.

In addition to the previously mentioned algorithms, the results relative to the minimum volume orthotope $\mathcal{O}^*(\Theta(t))$ are also reported in the examples. Such a set is computed by propagating the corresponding binding set $\mathcal{A}(t)$ in (12). The estimates provided by the standard RLS algorithm are also shown, along with the corresponding 99% confidence ellipsoids, which are compared to the approximating regions provided by the set membership approaches.

6.1 Example 1

Let us first reconsider the motivating example introduced in (1). By adopting the same settings as in Section 2, the considered methods are compared on a longer data set, with $N = 10000$. Results, averaged over 100 runs, are depicted in Figures 3-4, showing the volumes of the approximating regions and the number of constraints in the sets $\mathcal{A}(t)$ and $\mathcal{C}(t)$, respectively.

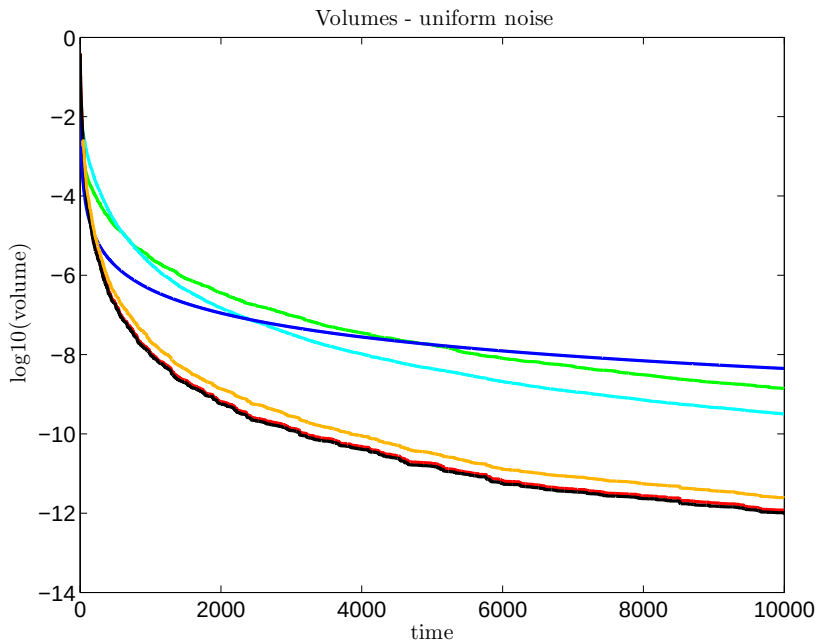


Figure 3: Example 1, uniform noise: volumes of approximating regions; ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}^*(\Theta(t))$ (black), $\mathcal{O}(t)$ from Procedure 1 (red), RLS (blue).

It can be observed that the volumes of the approximating orthotope $\mathcal{O}(t)$ are orders of magnitude smaller than that of ROBE, ROBP and RLS. The only comparable method is the RCE, which provides a volume about twice that obtained by the proposed algorithm. While the volume of $\mathcal{O}(t)$ is almost the same as the one of the minimum orthotope $\mathcal{O}^*(\Theta(t))$, the number of constraints that are propagated in the set $\mathcal{C}(t)$ is much lower than the number of constraints in the binding set $\mathcal{A}(t)$

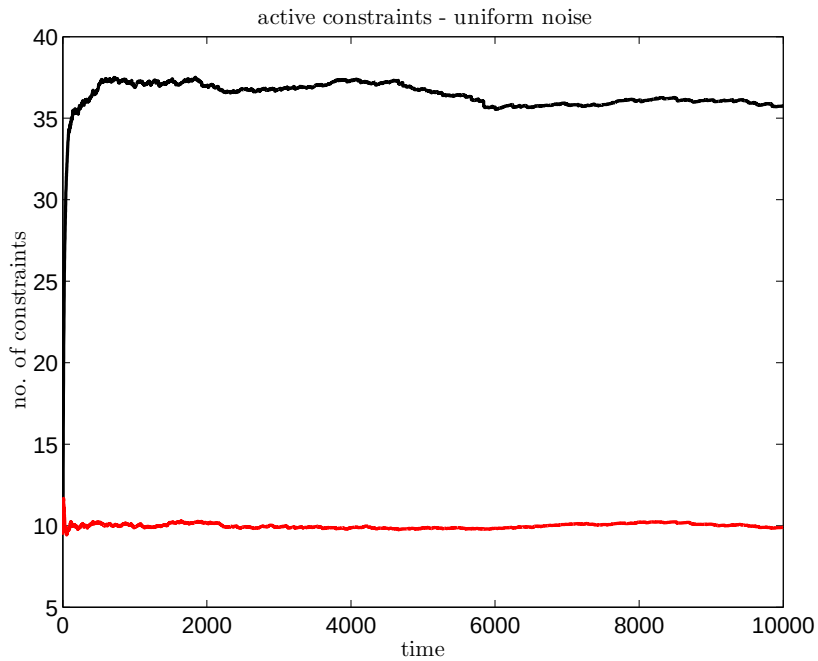


Figure 4: Example 1, uniform noise: number of constraints in the sets $\mathcal{A}(t)$ (black) and $\mathcal{C}(t)$ (red).

and, even more importantly, almost constant throughout the whole simulation. This guarantees that the computational burden of the proposed technique does not increase over time. Moreover, the computation of $\mathcal{O}(t)$ required on average the solution of only one LP every 37 time samples, while that of $\mathcal{O}^*(\Theta(t))$ involved the solution of one LP every 7 time samples.

The same test has been repeated with a Gaussian noise signal $e(t)$ of variance σ^2, truncated within the interval $[-3\sigma, 3\sigma] = [-0.1, 0.1]$. Results are reported in Figures 5-6. It can be noticed that the performance of the proposed method is still very close to that of the minimum orthotope, and much better than that of the ROBE, ROBP and RCE algorithms. The better performance of the RLS algorithm is not surprising in this case, because it is the minimum variance estimator in the case of Gaussian noise. Another important observation concerns the constraint sets. The number of constraints in $\mathcal{C}(t)$, propagated by the proposed technique, is almost the same as in the uniform noise

case, and still pretty constant over time. On the other hand, the number of binding constraints in $\mathcal{A}(t)$ is significantly higher and undergoes remarkable variations over time. The increased difference in the computational burden is testified also by the average number of time steps during which one LP is solved, which is 40 for the proposed technique and only 2 for the computation of $\mathcal{O}^*(\Theta(t))$.

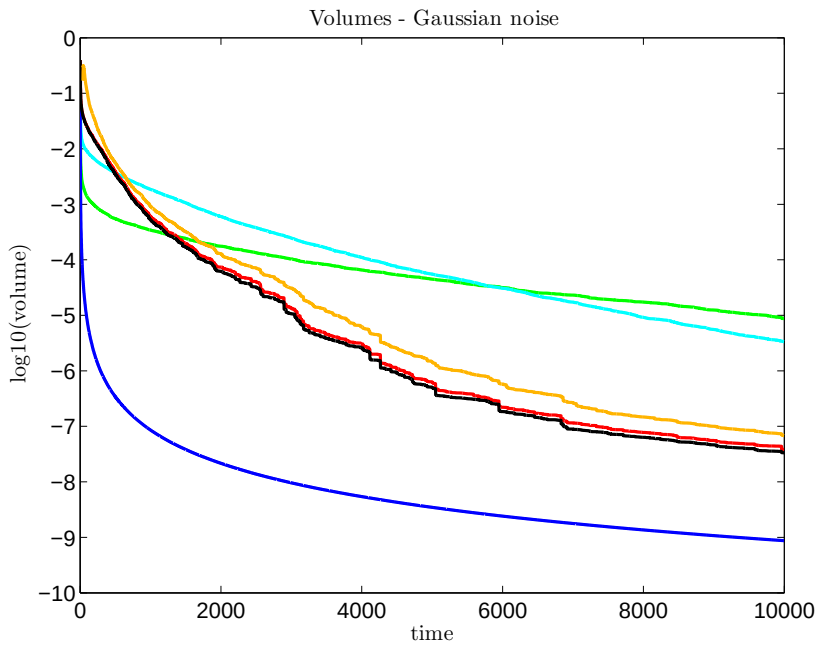


Figure 5: Example 1, Gaussian noise: volumes of approximating regions; ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}^*(\Theta(t))$ (black), $\mathcal{O}(t)$ from Procedure 1 (red), RLS (blue).

6.2 Example 2

The second case study concerns a sixth order ARX model with three pairs of lightly damped complex poles. The system equation is

$$\begin{aligned} &y(t) - 4.727y(t-1) + 9.899y(t-2) - 11.93y(t-3) + 8.824y(t-4) - 3.802y(t-5) + 0.7408y(t-6) \\ &\quad = 2.051u(t-1) - 5.486u(t-2) + 3.558u(t-3) + 2.899u(t-4) - 4.718u(t-5) + 1.711u(t-6) + e(t) \end{aligned}$$

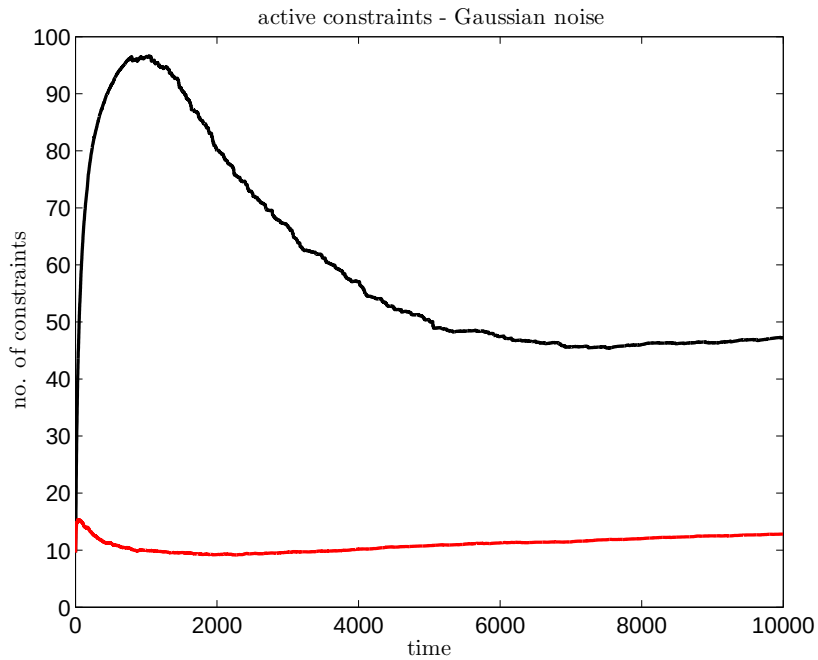


Figure 6: Example 1, Gaussian noise: number of constraints in the sets $\mathcal{A}(t)$ (black) and $\mathcal{C}(t)$ (red).

The number of parameters to be estimated is 12. Both the input $u(t)$ and the noise $e(t)$ are uniformly distributed in the interval $[-1, 1]$. Average results for 100 randomly generated input-output data sets, each one consisting of 10000 data points, are presented hereafter. Besides the volume of the approximating regions, also the following performance indexes are considered:

- the ℓ_2 relative parametric uncertainty

$$ru_2 = \max_{\vartheta \in \mathcal{R}} \frac{\|\overline{\theta} - \vartheta\|_2}{\|\overline{\theta}\|_2}; \tag{18}$$

- the ℓ_∞ relative parametric uncertainty

$$ru_\infty = \max_{i=1,\ldots,n} \max_{\vartheta \in \mathcal{R}} \frac{|\overline{\theta}_i - \vartheta_i|}{|\overline{\theta}_i|}. \tag{19}$$

Figure 7 reports the volumes of the considered techniques, Figure 8 shows the number of constraints in the sets $\mathcal{C}(t)$ and $\mathcal{A}(t)$, while Figure 9 presents the two performance indexes defined above.

As in Example 1, the new method outperforms the ROBE, ROBP and RCE algorithms, and the difference between its performance and that of the minimum bounding orthotope is negligible. A remarkable novelty with respect to the previous example is that the number of constraints in the set $\mathcal{A}(t)$ grows steadily as time passes, while the number of constraints in the set $\mathcal{C}(t)$ is much smaller and remains almost constant. It should be remarked that the proposed technique performs much better also with respect to the classical RLS algorithm, in terms of relative errors and parametric uncertainty, as shown by Figure 9. The parametric uncertainty in both ℓ_2 and ℓ_∞ norms is two orders of magnitude smaller with respect to those achieved by the ROBE and ROBP techniques, while it is about one half of that provided by the RCE method.

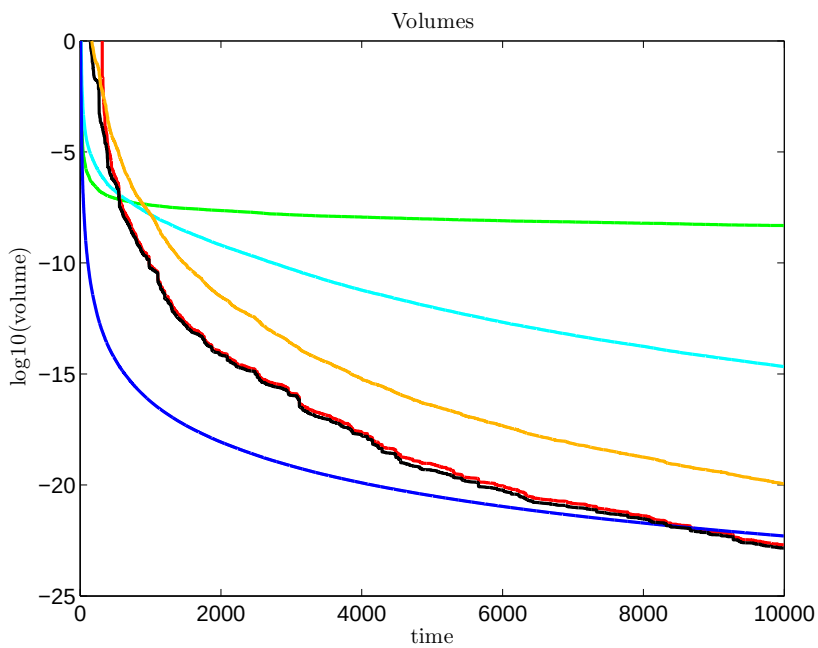


Figure 7: Example 2: volumes of approximating regions; ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}^*(\Theta(t))$ (black), $\mathcal{O}(t)$ from Procedure 1 (red), RLS (blue).

As long as the computational burden is concerned, Table 1 reports the average absolute[1] and

[1]Computations are performed under Matlab by using IBM ILOG CPLEX for MATLAB toolbox [41] to solve the

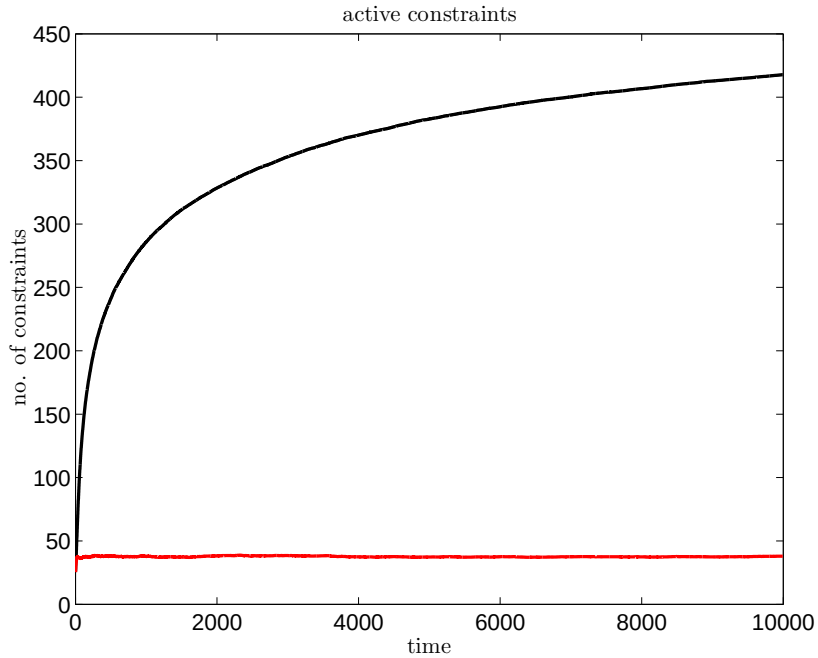
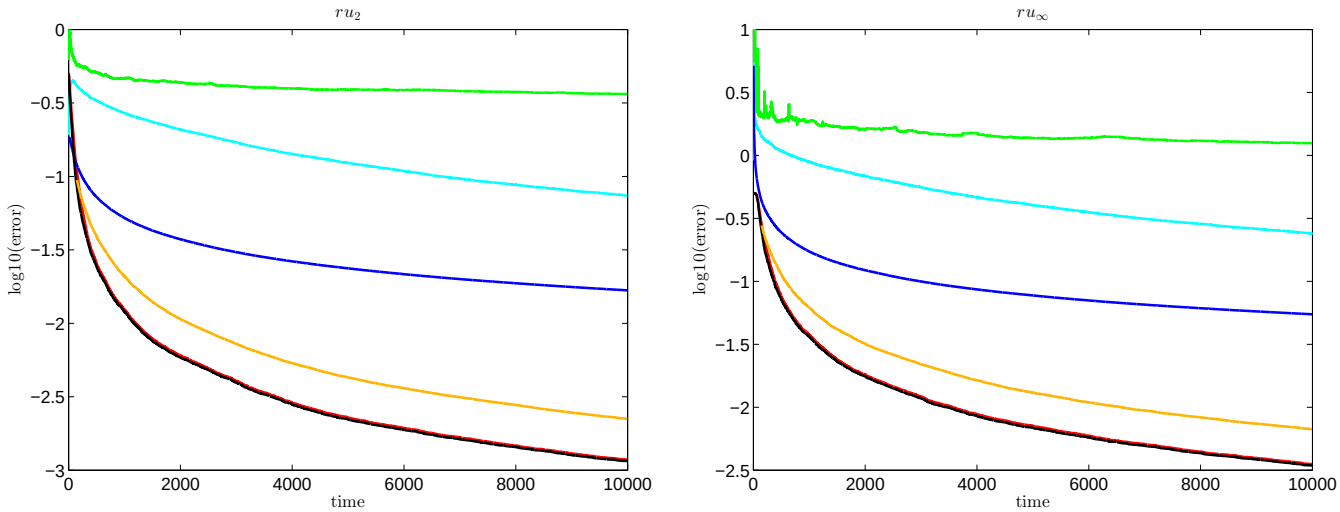


Figure 8: Example 2: number of constraints in the sets $\mathcal{A}(t)$ (black) and $\mathcal{C}(t)$ (red).



(a) (b)

Figure 9: Example 2: errors provided by ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}^*(\Theta(t))$ (black), $\mathcal{O}(t)$ from Procedure 1 (red), RLS (blue). (a) ru_2, (b) ru_∞.

	relative times	absolute times (ms)	#LPs/N
ROBE	1	0.065	-
ROBP	1.12	0.073	-
RCE	2.44	0.159	-
$\mathcal{O}(t)$	2.15	0.140	0.26
$\mathcal{O}^*(\Theta(t))$	59.14	3.852	2.89
RLS	0.59	0.038	-

Table 1: Example 2: Computational times and number of solved LPs per time step.

relative times per iteration required by each technique. Relative times are normalized to the time required by a single iteration of the ellipsoidal algorithm. It can be observed that the new method requires a time in the same order of other recursive algorithms, while a much higher computational burden is required by algorithm computing $\mathcal{O}^*(\Theta(t))$. This is due to the much larger average number of solved LPs per time step and due to the fact that the solved LPs contain a large number of constraints. Simulations with Gaussian noise realizations show the same features as those reported in Example 1.

6.3 Example 3

The aim of this last example is to analyze how the proposed method scales with the dimension of the parameter vector. Consider the FIR model class

$$y(t) = \sum_{i=1}^{n} \theta_i u(t-i) + e(t) \tag{20}$$

LPs, on an Intel Core i7-3770 at 3.40 GHz.

where the impulse response samples θ_i are the parameters to be identified and $e(t)$ satisfies (8). For different values of n, parameter vectors θ have been randomly selected, and for each one of them an identification experiment has been performed. In each experiment, the input $u(t)$ is uniformly distributed in $[-1, 1]$, $e(t)$ is uniformly distributed in $[-\delta, \delta]$ and δ has been chosen so that $\delta = 0.1\|y_{nf}(t)\|_\infty$, where $y_{nf}(t) = y(t) - e(t)$ is the noise-free output. For each value of n, average results over 100 FIR models and identification experiments, each one consisting of 20000 data points, are reported. Figure 10 compares the nth root of the final volume of the orthotope provided by the proposed technique to those obtained by the ROBE, ROBP, RCE, and by the 99% confidence ellipsoid of the RLS, for different values of n. It is apparent that the proposed technique provides a much better approximation of the feasible parameter set, not only with respect to the guaranteed set membership methods, but also to the probabilistic approximation given by the RLS confidence region. The same conclusion can be drawn from the final values of ru_2 and ru_∞, which are reported in Figures 11. In Fig. 12, the average number of constraints processed by the proposed technique for the case $n = 20$ is reported. As in Example 2, after a short transient, such a number remains approximately constant over time, settling to a value around 80: a much smaller value w.r.t. the number of constraints propagated by the RCE algorithm which is $2n^2 = 800$.

Another simulation study has been performed in the same setting, by generating the noise signal $e(t)$ from a Gaussian distribution of variance σ^2, truncated within the interval $[-3\sigma, 3\sigma]$, and σ chosen so that $3\sigma = 0.1\|y_{nf}(t)\|_\infty$. Results are shown in Figures 13-14. The conclusions that can be drawn are pretty much the same as in the case of uniform noise. The volumes and relative errors provided by the ROBE, ROBP and RCE approaches are once again much larger than those of the proposed technique. Another feature of the proposed method is its robustness with respect to different noise

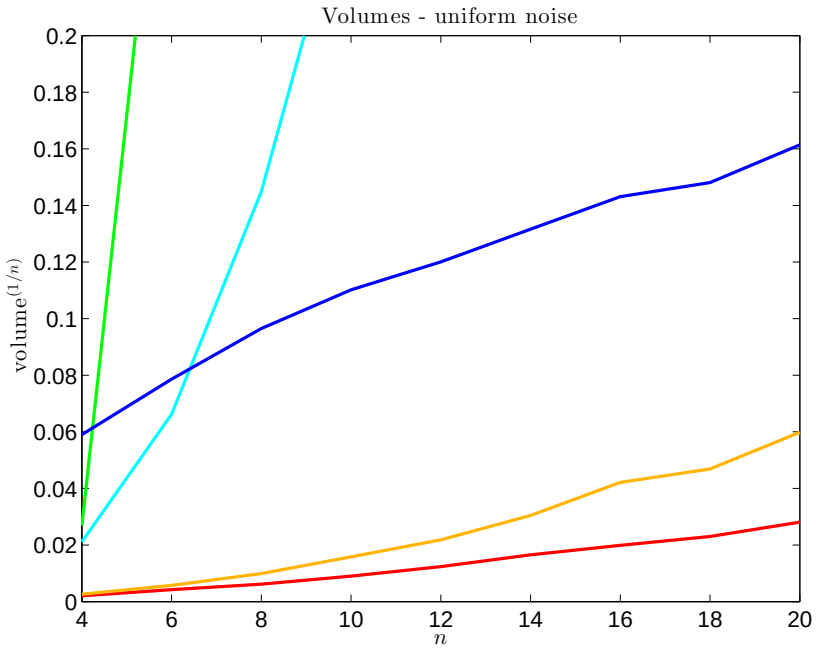
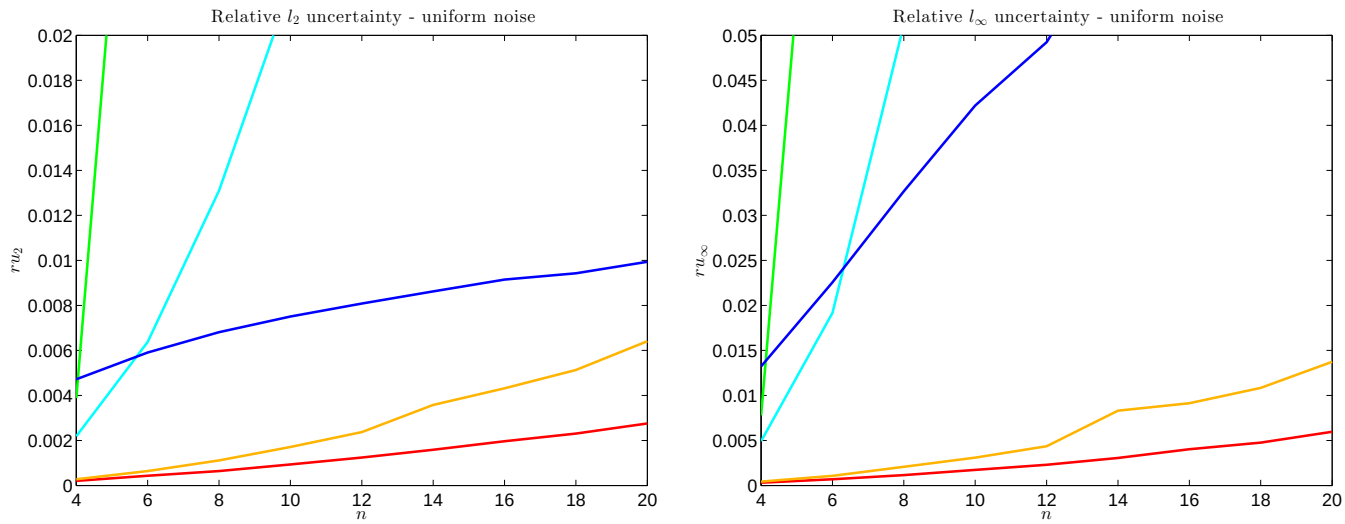


Figure 10: Example 3, uniform noise: nth root of the final volume of approximating regions as a function of the FIR model order n; ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}(t)$ (red), RLS (blue).



(a) (b)

Figure 11: Example 3, uniform noise: final errors provided by ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}(t)$ (red), RLS (blue) for different FIR model orders n. (a) ru_2, (b) ru_∞.

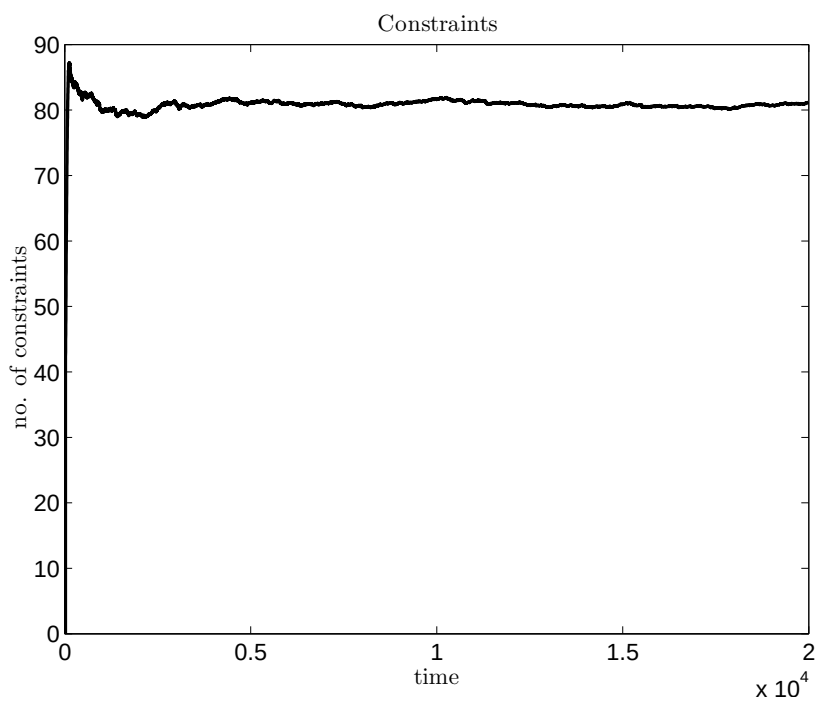


Figure 12: Example 3, uniform noise: average number of constraints in the set $\mathcal{C}(t)$ for $n = 20$.

realizations. For instance, the average ratio between the final ru_∞ of the RCE and that of the proposed technique is about 2, but it grows up to 80 for some realizations.

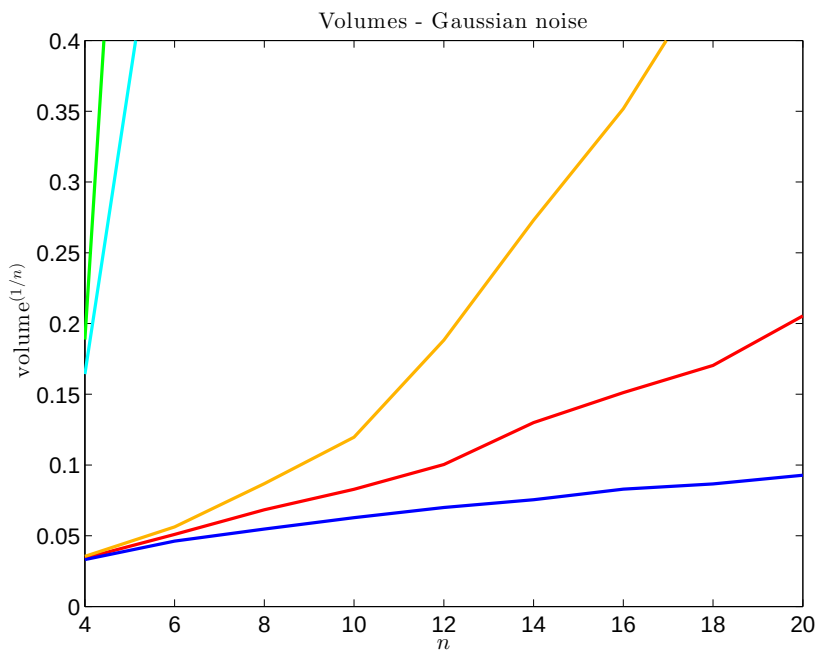
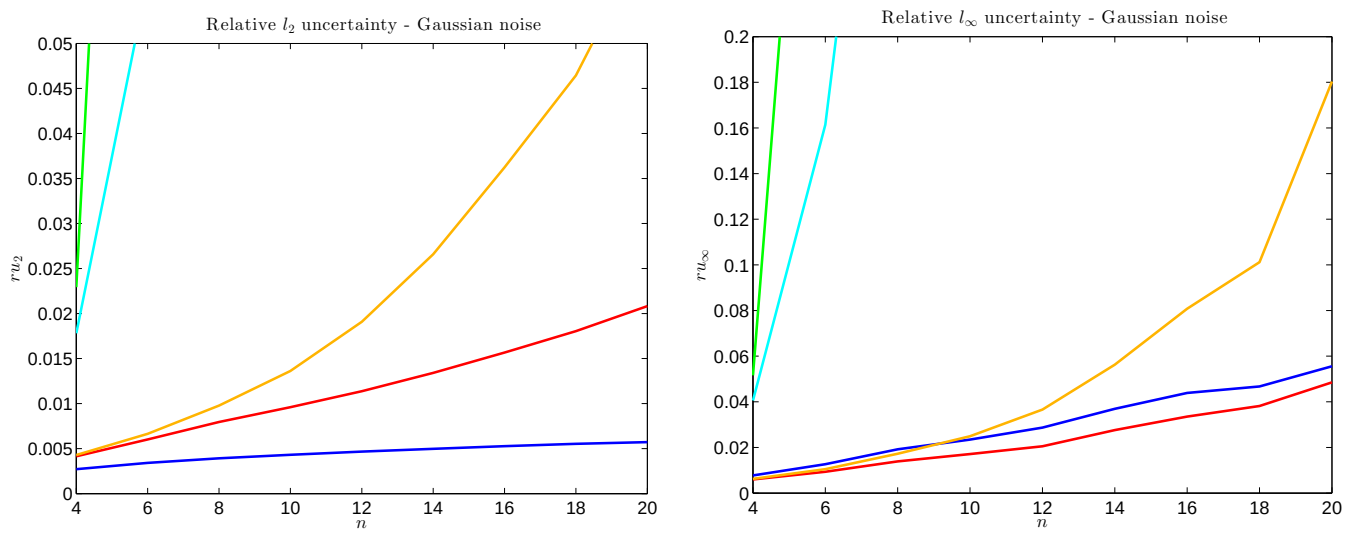


Figure 13: Example 3, Gaussian noise: nth root of the final volume of approximating regions as a function of the FIR model order n; ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}(t)$ (red), RLS (blue).

The proposed algorithm and other approximating techniques have been compared also for different values of the ℓ_∞ signal-to-noise ratio (SNR), defined as $\|y_{nf}\|_\infty/\|e\|_\infty$. In particular, the case of FIR models of order 10 has been analyzed, assuming uniformly distributed noise. In Table 2, the final volumes (after 20000 samples) of the sets provided by the different methods are reported. It can be noticed that the proposed algorithm outperforms the other techniques for any value of SNR.

As long as the computational burden is concerned, Figure 15 reports the average times per iteration, normalized to the average iteration time of the ellipsoidal algorithm, for both the uniform



(a) (b)

Figure 14: Example 3, Gaussian noise: final errors provided by ROBP (green), ROBE (cyan), RCE (yellow), $\mathcal{O}(t)$ (red), RLS (blue) for different FIR model orders n. (a) ru_2, (b) ru_∞.

and the Gaussian noise tests[2]. It can be observed that the time required by the proposed technique grows with the number of parameters to be estimated, up to approximately 4 times the average iteration time of the ellipsoidal algorithm for $n = 20$. This is basically due to a similar trend in the average number of LPs to be solved. Notice that the RCE technique is always more computationally demanding, due to a much larger number of orthotope updates.

A deeper investigation of the computational load during the time horizon of the simulations, suggests that a significant amount of the required computations is performed during the first part of the simulations. This is due to the fact that initially it is more likely that the measurement set $\mathcal{S}(t+1)$ introduces new active constraints into the FPS. In fact, the average iteration time of

[2]The average computational time of the ROBE algorithm for each time step for $n = 20$ is 0.075 and 0.070 milliseconds for uniform and Gaussian noise, respectively.

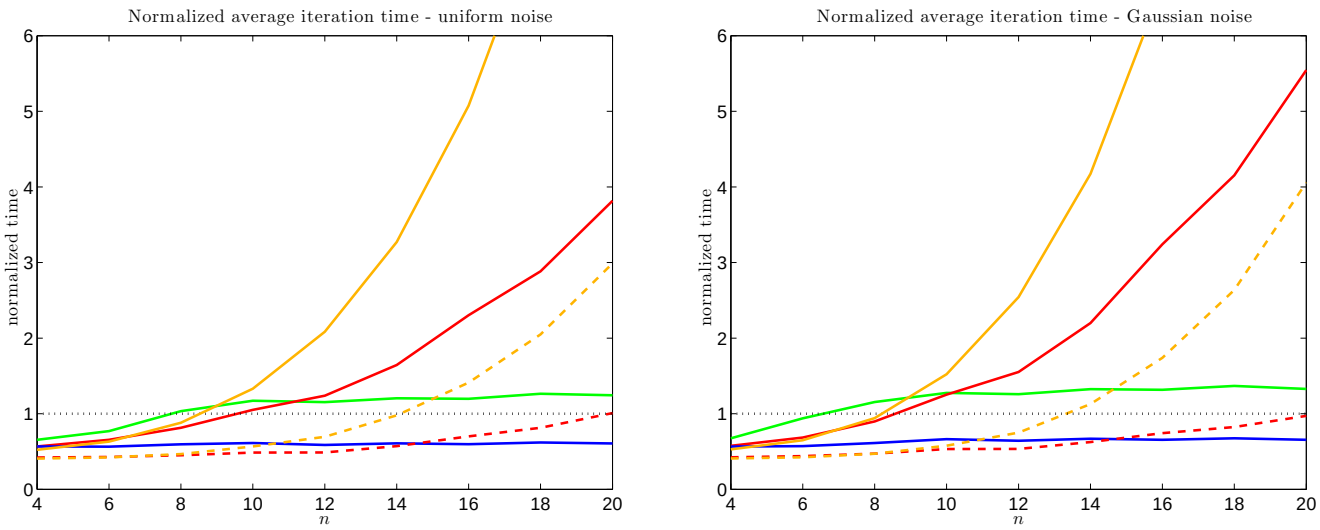
SNR	1	0.1	0.01	0.001
ROBE	-0.73	-8.46	-18.38	-28.38
ROBP	2.99	2.40	-9.83	-19.77
RCE	-10.56	-20.72	-30.61	-40.74
$\mathcal{O}(t)$	-12.85	-22.95	-32.88	-42.96
RLS	-2.55	-12.19	-22.13	-32.09

Table 2: Example 3: log-volumes for different values of the ℓ_∞ SNR, for $n = 10$.

the proposed technique in the second part of the simulations (last 10000 samples) turns out to be smaller than that of the ellipsoidal algorithm up to $n = 20$, and even smaller than that of the RLS for $n \leq 14$ (dashed curves in Figure 15). This means that when the estimation procedure reaches its "steady state", only few measurements contribute to reduce the approximating orthotope, and hence the number of LPs to be solved is quite small. A similar characteristic is observed also for the RCE algorithm, whose computational times however grow much faster with respect to the proposed technique, due to the much larger average number of orthotope updates.

7 Conclusions

This work is a first contribution in the direction of bridging the gap between the online and offline set membership parameter estimation techniques, in terms of both quality of the estimates and required computational burden. A new technique for computing online an orthotopic approximation of a polytopic feasible parameter set has been presented. The proposed approach exploits the efficiency of linear programming solvers, thanks to a suitable selection of the constraints to be propagated.



(a) (b)

Figure 15: Example 3, normalized average times per iteration: ROBP (green), RCE (yellow), proposed technique (solid red), RLS (blue), RCE (dashed yellow) and proposed technique (dashed red) after t=10000. (a) Uniform noise, (b) Gaussian noise.

The simulation studies show that it is sufficient to propagate a very limited number of constraints to achieve an approximation almost indistinguishable from the minimum orthotope containing the FPS, computed offline. Another important feature is that the number of propagated constraints and LPs to be solved remains approximately constant throughout the simulation tests, which makes the proposed technique feasible for online implementation, even in the presence of severe bandwidth limitations. The performance in terms of volumes of the approximating regions, nominal parametric errors and parametric uncertainty turns out to be remarkably better than those of recursive set approximation techniques available in the literature, at the price of a negligible increase of the computational burden.

The constraint selection approach proposed in this paper appears to be promising also in other estimation problems. Preliminary results on recursive identification of slowly time-varying systems have been presented in [42]. Current research activity is dedicated to the application of this approach to set membership state estimation problems.

References

[1] Ljung L. *System Identification: Theory for the User, 2nd ed.* Prentice Hall: Upper Saddle River, NJ, 1999.

[2] Åström KJ, Wittenmark B. *Adaptive control.* Courier Corporation, 2013.

[3] Isermann R. *Fault-diagnosis systems: an introduction from fault detection to fault tolerance.* Springer Science & Business Media, 2006.

[4] Bertsekas DP, Rhodes IB. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control* 1971; **16**:117–128.

[5] Schweppe FC. *Uncertain Dynamic Systems*. Prentice Hall: Englewood Cliffs, NJ, 1973.

[6] Kurzhanski A. Identification - a theory of guaranteed estimates. *From Data to Model*, Willems J (ed.). Springer Berlin, 1989; 135–214.

[7] Walter E, Piet-Lahanier H. Estimation of parameter bounds from bounded-error data: a survey. *Mathematics in Computer and Simulation* 1990; **32**:449–468.

[8] Milanese M, Vicino A. Estimation theory for nonlinear models and set membership uncertainty. *Automatica* 1991; **27**:403–408.

[9] Norton JP ((ed.)). *Special issue on: Bounded-error estimation - Part 1*, *International Journal of Adaptive Control and Signal Processing*, vol. 8(1), 1994.

[10] Norton JP ((ed.)). *Special issue on: Bounded-error estimation - Part 2*, *International Journal of Adaptive Control and Signal Processing*, vol. 9(1), 1995.

[11] Milanese M, Norton JP, Piet-Lahanier H, Walter E. *Bounding Approaches to System Identification*. Plenum Press: New York, 1996.

[12] Chernousko F, Polyak BT ((eds.)). *Special Issue on: The set membership modelling of uncertainties in dynamical systems*, *Mathematical and Computer Modelling of Dynamical Systems*, vol. 11(2), 2005.

[13] Veres SM ((ed.)). *Special issue on: Bounding methods for state and parameter estimation*, *International Journal of Adaptive Control and Signal Processing*, vol. 25(3), 2011.

[14] Cheng Y, Wang Y, Sznaier M, Ozay N, Lagoa C. A convex optimization approach to model (in)validation of switched arx systems with unknown switches. *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 2012; 6284–6290.

[15] Ozay N, Sznaier M, Lagoa C. Convex certificates for model (in)validation of switched affine systems with unknown switches. *IEEE Transactions on Automatic Control* 2014; **59**(11):2921–2932.

[16] Ozay N, Lagoa C, Sznaier M. Set membership identification of switched linear systems with known number of subsystems. *Automatica* 2015; **51**:180 – 191.

[17] Cerone V, Piga D, Regruto D. Bounded error identification of Hammerstein systems through sparse polynomial optimization. *Automatica* 2012; **48**(10):2693 – 2698.

[18] Cerone V, Piga D, Regruto D. Computational load reduction in bounded error identification of Hammerstein systems. *IEEE Transactions on Automatic Control* 2013; **58**(5):1317–1322.

[19] Cerone V, Lasserre JB, Piga D, Regruto D. A unified framework for solving a general class of conditional and robust set-membership estimation problems. *IEEE Transactions on Automatic Control* 2014; **59**(11):2897–2909.

[20] Casini M, Garulli A, Vicino A. Input design in worst-case system identification with quantized measurements. *Automatica* 2012; **48**(12):2997–3007.

[21] Casini M, Garulli A, Vicino A. Feasible parameter set approximation for linear models with bounded uncertain regressors. *IEEE Transactions on Automatic Control* 2014; **59**(11):2910–2920.

[22] Cerone V, Piga D, Regruto D. A convex relaxation approach to set-membership identification of LPV systems. *Automatica* 2013; **49**(9):2853 – 2859.

[23] Pshenichnyi BN, Pokotilo VG. A minimax approach to the estimation of linear regression parameters. *Eng. Cybernetics* 1983; :77–85.

[24] Belforte G, Tay TT. Two new estimation algorithms for linear models with unknown but bounded measurement noise. *IEEE Transactions on Automatic Control* 1993; **38**(8):1273–1279.

[25] Cerone V, Piga D, Regruto D. Improved parameters bounds for set-membership EIV problems. *International Journal of Adaptive Control and Signal Processing* 2011; **25**(3):208–227.

[26] Fogel E, Huang F. On the value of information in system identification - bounded noise case. *Automatica* 1982; **18**:229–238.

[27] Belforte G, Bona B, Cerone V. Parameter estimation algorithms for a set-membership description of uncertainty. *Automatica* 1990; **26**:887–898.

[28] Deller J, Nayeri M, Liu M. Unifying the landmark developments in optimal bounding ellipsoid identification. *International journal of adaptive control and signal processing* 1994; **8**(1):43–60.

[29] Vicino A, Zappa G. Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Transactions on Automatic Control* 1996; **41**:774–785.

[30] Chisci L, Garulli A, Vicino A, Zappa G. Block recursive parallelotopic bounding in set membership identification. *Automatica* 1998; **34**(1):15–22.

[31] Kostousova E. State estimation for dynamic systems via parallelotopes optimization and parallel computations. *Optimization Methods and Software* 1998; **9**(4):269–306.

[32] Bravo JM, Alamo T, Camacho EF. Bounded error identification of systems with time-varying parameters. *IEEE Transactions on Automatic Control* 2006; **51**(7):1144–1150.

[33] Blesa J, Puig V, Saludes J. Identification for passive robust fault detection using zonotope-based set-membership approaches. *International Journal of Adaptive Control and Signal Processing* 2011; **25**(9):788–812.

[34] Casini M, Garulli A, Vicino A. On worst-case approximation of feasible system sets via orthonormal basis functions. *IEEE Transactions on Automatic Control* 2003; **48**(1):96–101.

[35] Ferreres G, M'Saad M. Estimation of output error models in the presence of unknown but bounded disturbances. *International Journal of Adaptive Control and Signal Processing* 1997; **11**(2):115–140.

[36] Karmarkar N. A new polynomial-time algorithm for linear programming. *Combinatorica* 1984; **4**(4):373–395.

[37] Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press: Cambridge (UK), 2004.

[38] Casini M, Garulli A, Vicino A. A constraint selection technique for recursive set membership identification. *Proc. of 19th IFAC World Congress*, Cape Town, South Africa, 2014; 1790–1795.

[39] Broman V, Shensa MJ. A compact algorithm for the intersection and approximation of n-dimensional polytopes. *Mathematics and Computers in Simulation* 1990; **32**:469–480.

[40] Blesa J, Puig V, Saludes J. Robust fault detection using polytope-based set-membership consistency test. *IET Control Theory Applications* 2012; **6**(12):1767–1777.

[41] IBM. IBM ILOG Cplex Optimizer. Online: http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/.

[42] Casini M, Garulli A, Vicino A. A constraint selection technique for set membership estimation of time-varying parameters. *Proc. of 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, 2014; 1029–1034.