Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

A constraint selection technique for recursive set membership identification

M. Casini, A. Garulli, A. Vicino

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

Abstract: Recursive approximation of the set of feasible parameters is a key problem in set membership identification. In this paper a new technique is presented, aimed at recursively computing an outer orthotopic approximation of polytopic feasible parameter sets. The main idea is to exploit the concept of binding constraints in linear programs, in order to select a limited number of constraints providing a good approximation of the exact feasible set. Numerical tests demonstrate that the proposed technique outperforms existing recursive approximation algorithms, with a limited increase of the required computational burden.

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

1. INTRODUCTION

Set membership identification deals with the estimation of models in the presence of noise signals which are assumed to be unknown but bounded (UBB) in some norm. Although the basic idea dates back to the early works of Bertsekas and Rhodes [1971] and Schweppe [1973], it is from the 80s that a wide variety of techniques have been proposed to tackle a number of estimation problems in this framework (see e.g., Kurzhanski [1989], Walter and Piet-Lahanier [1990], Milanese and Vicino [1991], Chernousko and Polyak [2005] and references therein).

When the UBB noise assumption is enforced, a natural way to represent the uncertainty associated to the estimation problem is given by the so-called *Feasible Parameter Set (FPS)*, which is the set of all parameter vectors compatible with the available data and the UBB noise assumption. The shape and complexity of this 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.

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). Recursive FPS approximations are also in order when considering online estimation problems, which play a key role in a number of applications, including plant monitoring, fault detection, adaptive control and many others. A number of recursive approximation algorithms have been devised for linear regression models with ℓ_∞ bounded noise. Several classes of approximating regions have been considered in the literature, including orthotopes (Pshenichnyi and Pokotilo [1983]), ellipsoids (Fogel and Huang [1982], Belforte et al. [1990], Deller Jr. et al. [1993]), parallelotopes (Vicino and Zappa [1996], Chisci et al. [1998], Kostousova [1998]) and zonotopes (Bravo et al. [2006], Blesa et al. [2011]). While orthotopes provide very coarse approximations of the true feasible set, parallelotopes often return tighter estimates with respect to ellipsoids, especially when the number of available data is strongly limited. The computational complexity of these techniques turns out to be of the same order of the celebrated Recursive Least Squares (RLS) algorithm, which is the standard tool in statistical online estimation (Ljung [1999]). However, 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), being n the number of parameters to be estimated.

This work is an attempt to bridge the gap between the recursive and the batch approximations, both in terms of computational burden and quality of the estimates. The main idea is to exploit the availability of extremely efficient techniques for solving LPs, in order to provide an orthotopic approximation of the FPS as close as possible to the exact batch estimate, while retaining a computational burden similar to that of the recursive ellipsoidal and parallelotopic approximations mentioned above. To do this, a constraint selection technique based on the concept of *binding constraints* is proposed. It turns out that the proposed technique is able to provide very good approximations of the batch orthotope, by keeping track of a very small number of constraints and by solving a number of LPs which is orders of magnitude smaller than those required to compute the exact minimum orthotope at each time sample.

The paper is organized as follows. The problem formulation is presented in Section 2, along with basic material on linear programs. The constraint selection technique is illustrated in Section 3, while Section 4 reports several numerical tests which compare the proposed approach with existing recursive identification algorithms. Finally, concluding remarks are given in Section 5.

2. PROBLEM FORMULATION AND PRELIMINARIES

2.1 Problem formulation

Consider the linear regression model

$$y(t) = \varphi^T(t)\theta + e(t) \tag{1}$$

where $\theta \in \mathbb{R}^n$ is the vector of parameters to be estimated, $\varphi(t)$ is a vector containing past values of the system input and output signals, and $e(t)$ is an unknown-but-bounded noise signal such that

$$|e(t)| \leq \delta, \quad \forall t. \tag{2}$$

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

$$\begin{aligned}\Theta(t) &= \{\theta \in \Theta_0 : |y(k) - \varphi^T(k)\theta| \leq \delta, \quad k = 1, \dots, t\} \\ &= \bigcap_{k=1}^{t} \mathcal{S}(k)\end{aligned} \tag{3}$$

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. Being $\Theta(t)$ a polytope in the parameter space, the computation of the minimum axis-aligned box (orthotope) containing $\Theta(t)$ requires to solve $2n$ LPs with $2t$ constraints. Clearly, this is a feasible approach for a batch parameter estimation in which the number of available measurements N is fixed and limited.

When dealing with on line parameter estimation, recursive approaches for the approximation of the feasible set $\Theta(t)$ have to be employed. In the literature, a number of recursive approximation techniques have been proposed based on the following scheme

$$\mathcal{R}(t+1) \supseteq \mathcal{R}(t) \bigcap \mathcal{S}(t+1) \tag{4}$$

where $\mathcal{R}(t)$ is a set belonging to a predefined class of approximating regions $\mathcal{R}$, containing the true feasible set $\Theta(t)$ at a generic time t. By applying the above scheme, starting from a set $\mathcal{R}(0) \supseteq \Theta_0$, one has that $\Theta(t) \subseteq \mathcal{R}(t)$, $\forall t$. The techniques proposed in the literature differ by the class of approximating regions $\mathcal{R}$ (orthotopes, ellispoids, parallelotopes, zonotopes, etc.) and by the criterion used to compute $\mathcal{R}(t+1)$ in (4), the most popular one being the minimization of the volume of $\mathcal{R}(t+1)$.

2.2 Linear programming facts

In the following, we recall some basic facts about LPs which will be exploited in the paper. Consider the LP

$$\begin{array}{c}\max \;\; c^T x \\ \text{s.t.} \\ Ax \leq b\end{array} \;. \tag{5}$$

Let us denote by $\mathcal{X} = \{x | \; Ax \leq b\}$ the *constraint set*, and by $\Xi = \{x \in \mathcal{X} | \; x = \arg\max c^T x\}$ the *solution set* of the LP (5). Let a_i^T denote the ith row of matrix A. The constraint $a_i^T x \leq b_i$ is an *active constraint* at a feasible point $x \in \mathcal{X}$, if $a_i^T x = b_i$. The constraint $a_i^T x \leq b_i$ is a *binding constraint* of the LP (5) if there exists $x^* \in \Xi$ such that $a_i^T x^* = b_i$. Let $\mathcal{I}$ be the set of indexes i such that $a_i^T x \leq b_i$ is a binding constraint. Then, the set $\mathcal{A} : \{x | \; a_i^T x \leq b_i, \; i \in \mathcal{I}\}$ is referred to as the *binding set* of the LP (clearly $\mathcal{X} \subseteq \mathcal{A}$).

The following properties stem directly from the above definitions.

Proposition 1. Consider the LP

$$\begin{array}{c}\max \;\; c^T x \\ \text{s.t.} \\ x \in \mathcal{A}\end{array} \tag{6}$$

where $\mathcal{A}$ is the binding set of the LP (5). The LPs (5) and (6) have the same solution and the same solution set Ξ.

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

$$\begin{array}{c}\max \;\; c^T x \\ \text{s.t.} \\ x \in \mathcal{A} \bigcap \mathcal{H}\end{array} \;\geq\; \begin{array}{c}\max \;\; c^T x \\ \text{s.t.} \\ x \in \mathcal{X} \bigcap \mathcal{H}\end{array} \;. \tag{7}$$

Proposition 1 implies that no conservatism is introduced if the constraint set is replaced by the binding set. However, conservatism may arise if a new constraint $\mathcal{H}$ is added, as stated by Proposition 2.

3. CONSTRAINT SELECTION TECHNIQUE

In this section, the constraint selection technique proposed for the recursive approximation of the feasible set is presented. The class of approximating regions $\mathcal{R}$ is that of *orthotopes*, or axis-aligned boxes, defined as

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

where $\overline{\theta}, d, w \in \mathbb{R}^n$, $d_i \geq 0$, $i = 1, \dots, n$, and diag(d) is a diagonal matrix with diagonal equal to d. The sets $\mathcal{F}_i = \{\theta \in \mathcal{O} : \; \theta_i = \overline{\theta}_i + d_i\}$ and $\mathcal{F}_{i+n} = \{\theta \in \mathcal{O} : \; \theta_i = \overline{\theta}_i - d_i\}$, for $i = 1, \dots, n$, are the $(n-1)$-dimensional *faces* of the orthotope.

Let the vectors e_i, $i = 1, \dots, n$ denote the columns of the identity matrix. The minimum orthotope containing a polytope Θ, denoted by $\mathcal{O}^*(\Theta)$, can be computed by solving the $2n$ LPs

$$\beta_i = \begin{array}{c}\max \;\; e_i^T \theta \\ \text{s.t.} \\ \theta \in \Theta\end{array}, \quad \beta_{i+n} = \begin{array}{c}\min \;\; e_i^T \theta \\ \text{s.t.} \\ \theta \in \Theta\end{array} \tag{8}$$

for $i = 1, \dots, n$. Then, the minimum orthotope containing Θ is given by $\mathcal{O}^*(\Theta) = \mathcal{O}(\overline{\theta}^*, d^*)$, where

$$\overline{\theta}_i^* = \frac{\beta_i + \beta_{i+n}}{2}, \quad d_i^* = \frac{\beta_i - \beta_{i+n}}{2}, \quad i = 1, \dots, n.$$

Let us now consider the feasible parameter set $\Theta(t)$, defined according to (3), and the corresponding minimum bounding orthotope $\mathcal{O}^*(\Theta(t))$, obtained by computing the solutions $\beta_i(t)$ of the $2n$ LPs (8). Let $\mathcal{A}_i(t)$, $i = 1, \dots, 2n$, be the binding sets of the $2n$ LPs and define

$$\mathcal{A}(t) = \bigcap_{i=1}^{2n} \mathcal{A}_i(t). \tag{9}$$

Let $\Xi_i(t)$ $i = 1, \dots, 2n$, be the solution sets of the $2n$ LPs (8) and let $v^{(i)}(t) \in \Xi_i(t)$. The following result holds.

Proposition 3.
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 $i = 1, \dots, 2n$, then

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

Remark 1. 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 has a facet 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, it can be shown that condition iii) in Proposition 3 is not only sufficient but also necessary, i.e. if $v^{(i)}(t) \notin \mathcal{S}(t+1)$, then $\beta_i(t+1) < \beta_i(t)$ for $1 \le i \le n$, and $\beta_i(t+1) > \beta_i(t)$ for $n+1 \le i \le 2n$.

The results in Proposition 3 suggest a strategy for selecting a subset of the constraints of the feasible parameter set $\Theta(t)$, and then compute the approximating orthotope only with respect to the selected constraints. In order to illustrate the idea, let us first suppose that at a generic time t the following elements be available:

- the set $\mathcal{A}(t)$ defined in (9);
- $2n$ elements $v^{(i)}(t)$, $i = 1,\ldots,2n$, such that $v^{(i)}(t)$ belongs to the solution set of the LP

$$\begin{array}{c} \max\,(\min)\ e_i^T\theta \\ \text{s.t.} \\ \theta \in \mathcal{A}(t) \end{array} . \tag{10}$$

According to item i) in Proposition 3, the 2n LPs (10) provide the optimal bounding orthotope $\mathcal{O}^*(\Theta(t))$, i.e., no conservatism is introduced by replacing $\Theta(t)$ by $\mathcal{A}(t)$, as long as an orthtotopic approximation of the feasible parameter set is concerned. When a new measurement is processed at time $t+1$, if one computes the minimum orthotope containing the polytope defined by the intersection between the sets $\mathcal{A}(t)$ and $\mathcal{S}(t+1)$, conservatism may occur (see item ii)). Nevertheless, item iii) states that, if an element of the solution set of the ith LP belongs to $\mathcal{S}(t+1)$, then the new measurement will 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 (item iv)). Hence, a strategy for the recursive updating of an orthotope containing the feasible set $\Theta(t)$ can be devised as follows.

At a generic time t, assume that the following are available: (i) a set $\mathcal{C}(t) \supseteq \Theta(t)$, composed by a subset of the constraints of $\Theta(t)$; (ii) the minimum orthotope $\mathcal{O}(t) = \mathcal{O}(\overline{\theta}(t), d(t))$ such that $\mathcal{O}(t) \supseteq \mathcal{C}(t)$; (iii) $2n$ elements $v^{(i)}(t) \in \mathcal{C}(t) \bigcap \mathcal{F}_i(t)$, where $\mathcal{F}_i(t)$ are the faces of $\mathcal{O}(t)$, i.e., such that

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

for $i = 1,\ldots,n$. Let us denote by $\mathcal{C}_i(t)$ the constraints in $\mathcal{C}(t)$ that are active at any element of $\mathcal{C}(t)\bigcap \mathcal{F}_i(t)$.

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

$$v^{(i)}(t+1) = \begin{array}{c} \arg\max\,(\min) \quad e_i^T\theta \\ \text{s.t.} \\ \theta \in \mathcal{C}(t)\bigcap \mathcal{S}(t+1) \end{array} \tag{11}$$

and $\mathcal{C}_i(t+1) = \mathcal{A}_i(t+1)$, where $\mathcal{A}_i(t+1)$ is the binding set of the LP (11). Otherwise, if $v^{(i)}(t) \in \mathcal{S}(t+1)$, set $v^{(i)}(t+1) = v^{(i)}(t)$ and $\mathcal{C}_i(t+1) = \mathcal{C}_i(t)$. [1]

[1] In (11), the notation max(min) means that max holds for $i = 1,\ldots,n$, while min holds for $i = n+1,\ldots,2n$.

Step 2. For $i = 1,\ldots,n$, set $\mathcal{O}(t+1) = \mathcal{O}(\overline{\theta}(t+1), d(t+1))$, where

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

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

Step 3. Set

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

$t = t+1$, and go back to Step 1.

Notice that, by construction, $\mathcal{C}(t+1)$, $\mathcal{O}(t+1)$ and the elements $v_i^{(i)}(t+1)$ satisfy the same properties of their counterparts at time t, and hence the procedure can be iterated. Moreover, $\mathcal{O}(t+1) \subseteq \mathcal{O}(t)$, $\forall t$.

The rationale behind the above recursive procedure 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, as it will be shown by the numerical tests in Section 4, the procedure is able to provide a much tighter approximation with respect to standard recursive methods, while the number of LPs to be solved and of constraints to be kept track of, turns out to be very small with respect to those of the batch minimum orthotope $\mathcal{O}^*(\Theta(t))$.

In order to initialize the procedure, a possible choice is 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, and then pick $v^{(i)}(t_0)$, $i = 1,\ldots,2n$, from the solution sets of the $2n$ LPs and set $\mathcal{C}(t_0) = \mathcal{A}(t_0)$ (by doing this, the first t_0 measurements are used to initialize the procedure). As an alternative, one can choose as $\mathcal{O}(0)$ any orthotope containing the a priori set Θ_0, pick as $v^{(i)}(0)$, $i = 1,\ldots,2n$, any element on the ith face of $\mathcal{O}(0)$ and as $\mathcal{C}(0)$ the set of constraints defining $\Theta(0)$.

4. NUMERICAL TESTS

In this section, the constraint selection procedure is compared to classical recursive approaches, such as the minimum volume ellipsoidal approximation, firstly proposed in Fogel and Huang [1982], and the minimum volume parallelotopic approximation introduced in Vicino and Zappa [1996]. Moreover, we report the results relative to the minimum volume orthotope containing the feasible parameter set $\Theta(t)$, computed by propagating the corresponding binding set $\mathcal{A}(t)$. The estimates provided by the standard RLS algorithm are also considered, along with the corresponding 99% confidence ellipsoids, which are used for comparison with the estimates of the feasible parameter set provided by the set membership approaches.

In order to compare the results provided by the different techniques, besides the volume of the approximating regions $\mathcal{R}$, other performance indicators are considered:

- the relative error of the central parameter estimates
$$re_2 = \frac{\|\overline{\theta} - \theta\|_2}{\|\theta\|_2} \tag{12}$$
where $\overline{\theta}$ is the center of the approximating region $\mathcal{R}$, and θ is the true parameter vector;
- the relative parametric uncertainty
$$ru_2 = \max_{\vartheta \in \mathcal{R}} \frac{\|\overline{\theta} - \vartheta\|_2}{\|\overline{\theta}\|_2}; \tag{13}$$
- the maximum relative error
$$re_\infty = \max_{i=1,\ldots,n} \frac{|\overline{\theta}_i - \theta_i|}{|\theta_i|}; \tag{14}$$
- the maximum relative uncertainty
$$ru_\infty = \max_{i=1,\ldots,n} \max_{\vartheta \in \mathcal{R}} \frac{|\overline{\theta}_i - \vartheta_i|}{|\overline{\theta}_i|}. \tag{15}$$

4.1 Example 1

The first case study concerns an ARX model, generated by discretizing a fourth-order transfer function with two pairs of lightly damped complex conjugate poles. The system equation is $A(q)y(t) = B(q)u(t) + e(t)$ where

$$A(q) = 1 - 3.193q^{-1} + 4.360q^{-2} - 3.107q^{-3} + 0.954q^{-4}$$
$$B(q) = 0.072q^{-1} + 0.679q^{-2} + 0.618q^{-3} + 0.051q^{-4}$$

and $e(t)$ is unknown-but-bounded as in (2). Hence, the number of parameters to be estimated is $n = 8$.
We first consider identification experiments with 20000 data points, an input signal $u(t)$ uniformly distributed in $[-1, 1]$, and the noise $e(t)$ uniformly distributed in the interval $[-\delta, \delta]$, with $\delta = 0.2$. All results presented hereafter are averaged over 100 input/noise realizations. Figure 1 compares the volume (in logarithmic scale) of the orthotope provided by the proposed technique, with that of the exact minimum volume orthotope and of the recursive minimum volume ellipsoid and parallelotope. It can be observed that the new method outperforms classic recursive approaches, while the difference with the minimum bounding orthotope is negligible (in fact, the two plots are indistinguishable). It is worth noticing that this occurs, even if the number of constraints in the set $\mathcal{C}(t)$ is much smaller than the number of constraints in the active set $\mathcal{A}(t)$, as depicted in Figure 2.
Table 1 reports the performance indexes (12)-(15), computed at the final time sample of the identification experiment. The results confirm that the proposed methods outperforms the ellipsoid and parallelotopic recursive approximations, especially in terms of parametric uncertainties ru_2 and ru_∞: for example, the ellipsoidal algorithm has a 19.7% maximum relative uncertainty ru_∞, while that of the proposed method is less than 0.5%. On the other hand, the differences with respect to the true minimum orthotope $\mathcal{O}^*(\Theta(t))$ are always negligible. The performance of the classic RLS algorithm is similar to that of the recursive ellipsoidal approximation. As long as the computational burden is concerned, Table 2 reports the average times per iteration required by each technique, normalized to the time required by a single iteration of the ellipsoidal algorithm. It can be observed that the new method only requires a time 24% higher than the ellipsoidal algorithm, while a much higher computational burden is required by the exact algorithm. This is due to the much larger total number of LPs to be solved, reported in the rightmost column of Table 2.

Another identification experiment has been performed with the same type of input, but assuming a Gaussian

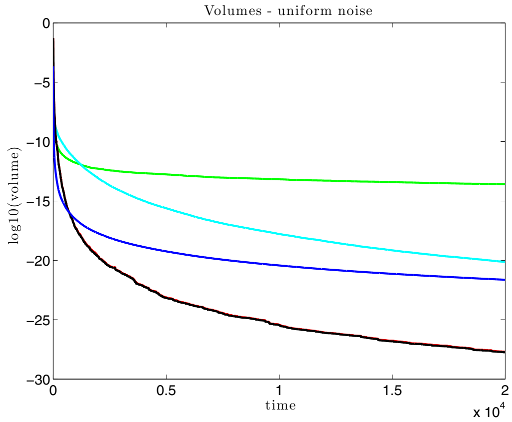

Fig. 1. Ex. 1, uniform noise: volumes of approximating regions: parallelotope (green), ellipsoid (cyan), minimum orthotope $\mathcal{O}^*(\Theta(t))$ (black), proposed approximating orthotope $\mathcal{O}(t)$ (red), RLS (blue).

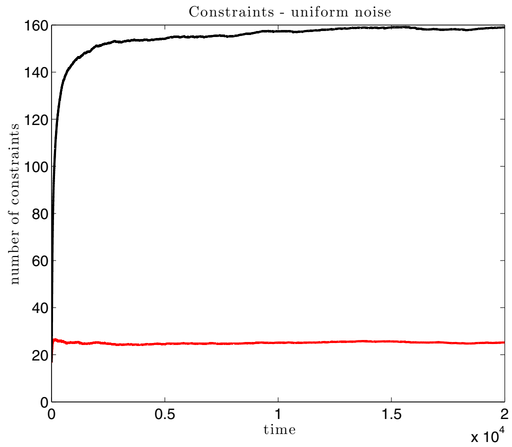

Fig. 2. Ex. 1, uniform noise: number of constraints in the sets $\mathcal{A}(t)$ (black) and $\mathcal{C}(t)$ (red).

	re_2	ru_2	re_∞	ru_∞
ellipsoid	2.32e-04	1.91e-03	1.38e-02	1.97e-01
parallelotope	2.72e-03	2.16e-02	2.72e-02	4.99e-01
$\mathcal{O}(t)$	1.97e-05	7.87e-05	1.09e-03	4.44e-03
$\mathcal{O}^*(\Theta(t))$	1.96e-05	7.74e-05	1.08e-03	4.36e-03
RLS	4.63e-04	1.13e-03	2.53e-02	1.25e-01

Table 1. Ex. 1, uniform noise: errors.

	relative times	no. of LPs
ellipsoid	1	-
parallelotope	1.12	-
$\mathcal{O}(t)$	1.24	1172
$\mathcal{O}^*(\Theta(t))$	9.47	12294
RLS	0.54	-

Table 2. Ex. 1, uniform noise: times and LPs.

noise signal $e(t)$ of variance σ^2, truncated within the interval $[-3\sigma, 3\sigma]$. Results are reported in Figures 3-4 and Tables 3-4. The main difference with respect to the previous case is the much larger number of active constraints in the set $\mathcal{A}(t)$, which results in a much higher computational burden of the procedure computing the exact bounding orthotope $\mathcal{O}^*(\Theta(t))$. Nevertheless, a similar increase is not observed in the number of constraints of $\mathcal{C}(t)$, and also the computational times and the number of LPs to be solved by the proposed technique remains pretty much the same as in the case of uniformly distributed noise. This demonstrates that, although the true feasible set is quite sensitive to the noise distribution, this does not affect significantly the proposed approximation technique. As long as the performance indicators are concerned, the proposed method still outperforms the ellipsoidal and parallelotope algorithms in terms of both nominal error and uncertainty. It can be noticed that the performance of the RLS algorithm is now better in terms of volume and comparable to that of the proposed technique for the uncertainty performances ru_2 and ru_∞ (although slightly worse in terms of nominal errors re_2 and re_∞). This is in accordance with the RLS being the minimum variance estimator in the case of Gaussian noise.

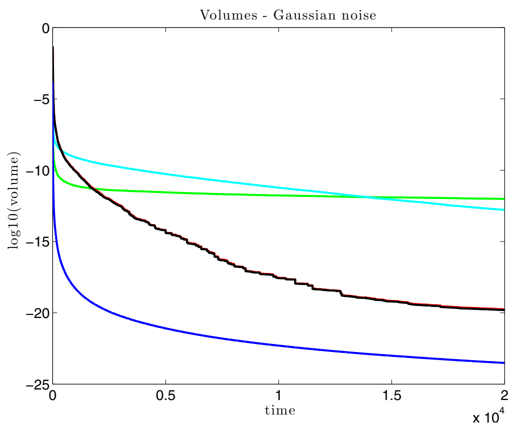

Fig. 3. Ex. 1, Gaussian noise: volumes of approximating regions: parallelotope (green), ellipsoid (cyan), minimum orthotope $\mathcal{O}^*(\Theta(t))$ (black), proposed approximating orthotope $\mathcal{O}(t)$ (red), RLS (blue).

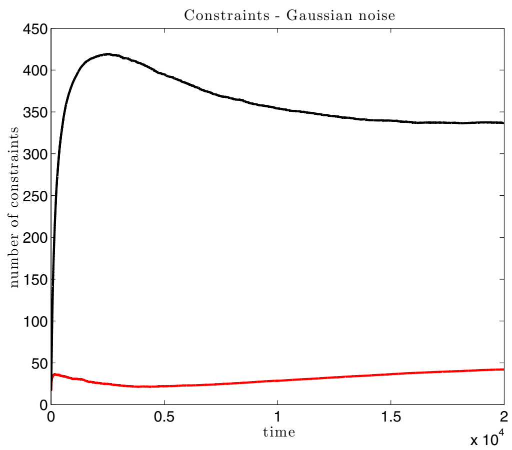

Fig. 4. Ex. 1, Gaussian noise: number of constraints in the sets $\mathcal{A}(t)$ (black) and $\mathcal{C}(t)$ (red).

	re_2	ru_2	re_∞	ru_∞
ellipsoid	2.14e-03	1.59e-02	1.27e-01	1.66e+00
parallelotope	6.00e-03	3.97e-02	4.59e-02	5.00e-01
$\mathcal{O}(t)$	8.47e-05	9.26e-04	4.33e-03	5.34e-02
$\mathcal{O}^*(\Theta(t))$	8.26e-05	9.12e-04	4.11e-03	5.24e-02
RLS	2.69e-04	6.62e-04	1.66e-02	7.43e-02

Table 3. Ex. 1, Gaussian noise: errors.

	relative times	no. of LPs
ellipsoid	1	-
parallelotope	1.26	-
$\mathcal{O}(t)$	1.39	1243
$\mathcal{O}^*(\Theta(t))$	47.37	35895
RLS	0.5624	-

Table 4. Ex. 1, Gaussian noise: times and LPs.

4.2 Example 2

In order to analyze how the proposed method scales with the dimension of the parameter vector, identification of the impulse response of FIR models of different lengths has been addressed. Consider the FIR model class

$$y(t) = \sum_{i=1}^{n} \theta_i u(t-i) + e(t) \qquad (16)$$

where the impulse response samples θ_i are the parameters to be identified and $e(t)$ satisfies (2). 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 which 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 the signal-to-noise ratio (SNR) in the ℓ_∞ norm is equal to 0.1. For each value of n, average results over 100 FIR models and identification experiments, each one consisting of 20000 data points, are evaluated.
In Table 5, numerical values of the nth root of the final volume, re_2 and ru_2 for different values of n, are reported and compared with those obtained by the recursive set membership ellipsoid and by the 99% confidence ellipsoid of the RLS. 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 ellipsoid, but also to the probabilistic approximation given by the RLS confidence region. A similar behavior has been observed for re_∞ and ru_∞.

		Uniform noise			Gaussian noise		
n	Algor.	vol$^{(1/n)}$	re_2	ru_2	vol$^{(1/n)}$	re_2	ru_2
4	$\mathcal{O}(t)$	9.9e-4	6.9e-5	2.0e-4	1.8e-2	1.9e-4	4.0e-3
	ellips.	1.1e-2	4.0e-4	2.1e-3	8.6e-2	2.9e-3	1.7e-2
	RLS	3.0e-2	2.3e-3	4.5e-3	1.7e-2	1.3e-3	2.6e-3
8	$\mathcal{O}(t)$	3.6e-3	1.7e-4	6.2e-4	3.7e-2	5.1e-4	7.4e-3
	ellips.	8.3e-2	1.8e-3	1.3e-2	6.4e-1	1.6e-2	9.7e-2
	RLS	5.5e-2	4.0e-3	6.6e-3	3.1e-2	2.1e-3	3.7e-3
12	$\mathcal{O}(t)$	6.5e-3	2.5e-4	1.2e-3	5.7e-2	1.2e-3	1.1e-2
	ellips.	2.2e-1	3.9e-3	3.1e-2	1.1e0	2.3e-2	1.7e-1
	RLS	6.6e-2	5.2e-3	7.9e-3	3.9e-2	3.0e-3	4.6e-3
16	$\mathcal{O}(t)$	1.0e-2	3.7e-4	1.9e-3	8.1e-2	1.9e-3	1.5e-2
	ellips.	4.1e-1	6.1e-3	5.5e-2	1.4e0	2.3e-2	2.0e-1
	RLS	7.8e-2	6.0e-3	9.0e-3	4.6e-2	3.4e-3	5.2e-3
20	$\mathcal{O}(t)$	1.5e-2	4.7e-4	2.7e-3	1.1e-1	3.2e-3	2.1e-2
	ellips.	6.1e-1	8.1e-3	8.0e-2	1.5e0	2.4e-2	2.3e-1
	RLS	8.7e-2	6.8e-3	9.9e-3	5.0e-2	3.9e-3	5.7e-3

Table 5. Ex. 2, uniform e Gaussian noise: errors.

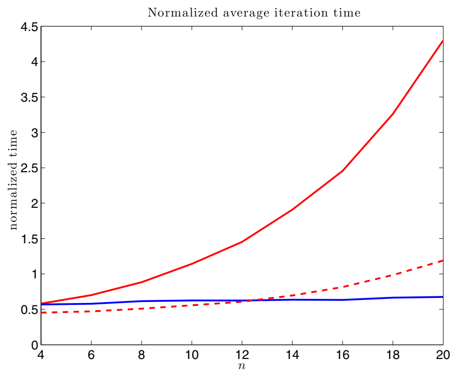

Fig. 5. Ex. 2, uniform noise: normalized average times per iteration: proposed technique (solid red), proposed technique after $t = 10000$ (dashed red), RLS (blue).

Figure 5 shows the average times per iteration, normalized to the average iteration time of the ellipsoidal algorithm. As expected, the time required by the proposed technique grows exponentially with the number of parameters to be estimated, up to approximately 4 times the average iteration time of the ellipsoidal algorithm for $n = 20$. However, the average iteration time of the proposed technique in the last 10000 time samples (dashed curve) turns out to be smaller than that of the ellipsoidal algorithm up to $n = 18$, and even smaller than that of the RLS for $n \leq 12$. This is due to the fact 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. In fact, a new measurement $y(t+1)$ is able to tighten the jth face of the current orthotope $\mathcal{O}(t)$ only if $v^{(j)} \notin \mathcal{S}(t+1)$, for some $j = 1, \ldots, 2n$ (see Proposition 3, item iii)), and this can be checked by computing only a single scalar product in $\mathbb{R}^n$. Hence, $2n$ scalar products are the only computations performed at the time samples in which the orthotope is not updated.

The same campaign of identification experiments has been repeated for a Gaussian noise of variance σ^2, truncated within the interval $[-3\sigma, 3\sigma]$, and σ chosen so that the ℓ_∞ SNR is equal to 0.1. Similar trends with respect to the tests with uniformly distributed noise have been observed (see Table 5), the main difference lying in that the volume of the RLS confidence ellipsoid is now slightly smaller than that of the orthotope, while both are orders of magnitude smaller than the volume of the set membership ellipsoid. The same occurs for the relative ℓ_2 uncertainty ru_2, while the orthotope provides a slightly more precise nominal estimate re_2 with respect to the RLS.

5. CONCLUSIONS

A new technique for computing a recursive approximation of a polytopic feasible parameter set has been presented. The proposed approach combines the efficiency of linear programming solvers with an appropriate selection of the constraints to be propagated. Numerical tests have shown that the computed approximation is almost indistinguishable from the exact minimum orthotope containing the feasible set, while it outperforms set membership algorithms based on the recursive optimization of approximating regions such as ellipsoids and parallelotopes. Moreover, the proposed method provides a valuable competitor to the statistical confidence regions provided by the celebrated RLS algorithm. Although the computational burden grows exponentially with the number of parameters to be estimated, it turns out to be of the same order of classic recursive approaches for systems of moderate size. The extension of the proposed technique to recursive identification of slowly time-varying systems and to set membership state estimation problems is the subject of ongoing research.

REFERENCES

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

D. P. Bertsekas and I. B. Rhodes. Recursive state estimation for a set-membership description of uncertainty. *IEEE Transactions on Automatic Control*, 16:117–128, 1971.

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

J. M. Bravo, T. Alamo, and E. F. Camacho. Bounded error identication of systems with time-varying parameters. *IEEE Transactions on Automatic Control*, 51(7):1144–1150, 2006.

F. Chernousko and B. T. Polyak. (Eds.), Special Issue on: The set membership modelling of uncertainties in dynamical systems. *Mathematical and Computer Modellimg of Dynamical Systems*, 11(2), 2005.

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

J. R. Deller Jr., M. Nayeri, and S. F. Odeh. Least-square identification with error bounds for real-time signal processing and control. *Proceedings of the IEEE*, 81:815–849, 1993.

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

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

A.B. Kurzhanski. Identification - a theory of guaranteed estimates. In J.C. Willems, editor, *From Data to Model*, pages 135–214. Springer Berlin, 1989.

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

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

B. N. Pshenichnyi and V. G. Pokotilo. A minimax approach to the estimation of linear regression parameters. *Eng. Cybernetics*, pages 77–85, 1983.

F. C. Schweppe. *Uncertain Dynamic Systems*. Prentice Hall, Englewood Cliffs, NJ, 1973.

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

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